Bankruptcy Prediction and its Advantages
Empirical Evidence from SMEs in the French Hospitality Industry

Author
Joseph Janer*

Academic Supervisor
Cédric Schneider

Master’s Thesis
Department of Economics
Copenhagen Business School
December, 2011

* The author can be contacted at: joseph.janer@gmail.com
No. of effective pages: 80; No. of characters with spaces: 185,957
Abstract

This study is about bankruptcy prediction modeling and explores the benefits from its application. Bankruptcies affect all stakeholders: from employees to regulators, investors or managers. Therefore, it is very interesting to understand the phenomenon that leads to bankrupt in order to take advantage of it.

The study begins with an exhaustive literature review with the purpose of understanding well the topic of bankruptcy prediction. Most of the models and techniques of bankruptcy prediction modeling up to this date are covered here. The main research questions that define this study are: (i) How to predict bankruptcies on a specific industry? (ii) How to attribute probabilities of bankruptcy and classes of risk to these predictions? (iii) How to determine the contributing variables to a predicted bankrupt and to benefit from it?

Linear discriminant analysis (LDA) method is used to answer these questions. Empirical evidence supports the developed model and study. The rate of good classification is equal to 86.36% of the holdout sample. Type I and II errors are in equivalent proportions after being rebalanced with a cut-off modification achieved by nonlinear programming optimization. Various testing of the model robustness are performed, such as logistic regression, which confirms the significance of the most of the explanatory variables. In order to refine the classification output of the model (either bankrupt or non-bankrupt firms), five classes of risks are developed – from the most to the least risky. In addition, probabilities of default and confidence intervals of the results are presented. Finally, a deeper examination of the results’ outputs is conducted and contributions from the different ratios that influence the model are analyzed.
# Contents

Abstract .......................................................................................................................................................... i

1. Introduction ............................................................................................................................................... 1

2. Research process ..................................................................................................................................... 3
   2.1. Motivations ....................................................................................................................................... 3
   2.2. Research Questions ...................................................................................................................... 3
   2.3. Limitations ....................................................................................................................................... 4
   2.4. Contributions .................................................................................................................................. 4

3. Literature review ..................................................................................................................................... 5
   3.1. Background and history ................................................................................................................ 5
   3.2. Earlier techniques ........................................................................................................................ 5
   3.3. Evolution of statistical techniques ............................................................................................... 6
   3.4. Alternative modeling techniques ............................................................................................... 10
   3.5. Evolution and empirical applications in France ........................................................................ 15
   3.6. Focused models ........................................................................................................................... 17

4. Theoretical model ................................................................................................................................... 19
   4.1. Geometrical approach ................................................................................................................ 19
   4.2. Probabilistic approach ............................................................................................................... 21

5. Data ...................................................................................................................................................... 27
   5.1. Data specificities ....................................................................................................................... 28
   5.2. Data reprocessing ...................................................................................................................... 33

6. Model development ............................................................................................................................... 37
   6.1. Selection of discriminant variables ........................................................................................... 37
   6.2. Combination of variables to form the discriminatory model ................................................... 41

7. Empirical results ..................................................................................................................................... 43
   7.1. Validation of a function estimate .............................................................................................. 43
   7.2. Validation of the final function ................................................................................................. 45
   7.3. Optimizing the cut-off value through nonlinear programming ........................................... 49
   7.4. Optimized results ....................................................................................................................... 52
1. Introduction

Companies are never protected against bankruptcy. Either in an economic expansion or in a recession, firms are likely to go bankrupt. An important competitive advantage is to understand the phenomenon that leads to bankrupt and to benefit from it.

The purpose of this study is to assess this issue. This is attempted by studying small and medium sized enterprises (SMEs) in the French Hospitality Industry. This study combines both theoretical and empirical interest. From a theoretical perspective this study applies well-known theories, and from an empirical perspective it provides elements for concrete utilization.

The particularity of the topic of firms’ bankruptcy is that it affects all stakeholders: employees, stockholders, managers, investors, and regulators. This study provides benefits for everyone interested to learn about modeling and bankruptcy prediction application.

Predicting bankruptcy is a difficult exercise and many challenges have to be faced. The first challenge starts with the selection of the technique to be used. For this reason, after initiating the research process, two sections are dedicated on the modeling techniques: one on the literature and the other – on the methodology. They are developed in order to determine the most appropriate technique to help answering the research questions raised. Mostly of the bankruptcy prediction techniques are covered in the literature review. Then, the specific theoretical methodology of the technique chosen – the Linear Discriminant Analysis, is presented.

Once the model technique is understood, appropriate data are gathered. However, data often need to be reprocessed according to appropriate techniques. This last step is very important for the rest of the study because the better the incoming data, the better the results of the study. This step is very long in the model development and can take up to 4/5th of the time dedicated to the study (Bardos, 2001). Thus, obtaining good quality data is a must.

Once data are gathered and reprocessed, they are ready and appropriate to be used in the modeling technique. The model development is composed of two steps: univariate and multivariate. First, overall selected ratios are tested. The most appropriate and discriminatory
ones are selected for the second phase. Second, ratios are assembled in a multivariate model according to their combined discriminatory abilities.

After data are combined, results from different models are interpreted and refined. Specifically, the results’ section is composed of four parts. First, each combination of ratios composing a model is tested on different estimates to select the best estimate from each model. Second, among the different models, the final model is selected. Third, results from this final model are adjusted by nonlinear programming. Lastly, following these adjustments, the final model results are explored.

The next step is to verify the robustness of the model. Different tests on major assumptions and overall goodness of fit are performed. For testing the assumptions, tests on multivariate normality, homoscedasticity and multicollinearity are performed. For the overall goodness of the model, a logistic regression is performed to test the significance of the variables included in the final model. Other tests, such as test on the equality of group means, the Eigenvalues and Wilk’s lambda, are performed.

The analysis of the developed model is further studied more in detail. Probabilities of bankruptcy according to different scores’ intervals are generated. The initial classification, opposing bankrupt to non-bankrupt firms, is refined into five classes of risk. In addition, uncertainty accompanying probabilities of bankruptcy and risk classes is modeled through confidence intervals.

Then, benefits from the model are further explored. Specific variables contributing to the scores are analyzed over time. For example, results from the industrial sector, as well as from particular situations encountered by firms, are analyzed. Finally, more general observations on the performed study are discussed.

The layout of the study is as follow: Section 2 covers the research process; Section 3 presents an exhaustive literature review of prior researches; Section 4 presents the theoretical model; Section 5 describes the data; Section 6 presents the development to the final model; Section 7 presents the empirical results; Section 8 presents the robustness analysis; Section 9 presents the analysis of the risk classes and posterior probabilities; Section 10 presents the analysis of contributions; Section 11 discusses the research study, and section 12 concludes.
2. **Research process**

2.1. **Motivations**

This subject was chosen because it allows working on both practical and theoretical aspects of a firm’s life. Bankruptcy’s study has recently become a hot topic due to the worldwide economic turmoil. This topic is very interesting and challenging. It concerns many actors of the business world and therefore, once achieved, the results should benefit the whole community. It is a good motivation to attempt to capture and understand the elements and reasons that lead to a corporate default. An additional motivation is to develop and implement a quantitative model to predict bankruptcies on SMEs that are the most present in the economic world\(^2\). Finally, one last motivation is, once the model developed, to benefit from it and to advance in understanding of bankruptcies.

2.2. **Research Questions**

To begin on this thesis, the initial intention was to study the risks of doing business with another firm that might go bankrupt. Therefore, the initial draft research question was: How to predict the bankruptcy of a company and to avoid default on payments? However, this angle of research was restricted to firms interacting with other firms.

As the subject of bankruptcy concerns all stakeholders, it is preferred to study the topic of bankruptcy prediction from a broader perspective, as on industry level, for example. In addition, it would be beneficial to take advantage of the bankruptcy prediction and to determine: (i) the probabilities of bankruptcy of a certain prediction, and (ii) the contributing variables to a predicted outcome.

Consequently, the final research questions that define this thesis are: (i) How to predict bankruptcies on a specific industry? (ii) How to attribute probabilities of bankruptcy and classes of risk to these predictions? (iii) How to determine the contributing variables to a predicted bankrupt and to benefit from it?

\(^2\) For example, with 23 million, SMEs in the EU represent 99% of businesses. Source: European Commission’s website, [http://ec.europa.eu/enterprise/policies/sme/index_en.htm](http://ec.europa.eu/enterprise/policies/sme/index_en.htm)
2.3. Limitations

In order to structure this study, three major limitations are set. The first one concerns the model. It is not possible before starting the literature review of bankruptcy prediction to determine and to identify a model to answer the questions raised. The model will be determined after studying the literature review. However, the model used will apply only quantitative data. Specifically, only financial statement data will be selected as they are available for everyone and should objectively contribute to answer the questions raised.

Second, this limitation concerns the economic impact of the thesis. This study will focus on SMEs³ because these firms are at the core of any industry and represent its biggest share. They are often well established in their business segment and should be less complex to analyze than multinationals or micro and start-up firms (Stili, 2002).

Finally, the last main limitation concerns the data. In order to satisfy the research questions raised, this study will focus on a specific industry – the French Hospitality Industry. In addition, detailed specificities and limitations of the data are explained in the data section.

2.4. Contributions

This study contributes in several aspects to this domain. It starts with an exhaustive literature review of the studies conducted on bankruptcy predictions up-to-date. The study applies a well-known methodology – the Linear Discriminant Analysis, to an unprecedented targeted population (specific niche of SMEs’ firms of the French hospitality industry). The advantage to focus the model on a specific industry allows tailoring it to the industry’s specific needs for better results. It provides different perspectives and additional possibilities of analysis and interpretation, such as: the use of nonlinear programming for cut-off optimization, various tests on the model’s robustness (including a logistic regression), and analysis of risks and variables contributing to the score output.

---
³ Note that the term SMEs such as used in this study refers to a subjective category of companies, which is detailed in the data section
3. Literature review

3.1. Background and history

The analysis of corporate distress traces its history back to two centuries ago (E. I. Altman & Edith Hotchkiss, 2006). At first, potential corporate distresses were assessed based on some qualitative information, which were very subjective. In particular, four references were mostly used, such as: (i) the capacity of the manager in charge of the project or company, (ii) the fact that the manager had an important financial involvement in the company as a financial guarantee, (iii) the project and the industry in itself, and (iv) the fact that the firm possessed assets or collateral to back-up in case of a bad situation. Surprisingly, these recommendations could still be considered in many existing investment decisions.

Later, early in the 20th century, the analysis of companies’ financial conditions has moved forward to the analysis of financial statement data, more particularly, to the univariate ratio analysis. It is also interesting to mention that during this period were found some of the most successful contemporary companies in the analysis of the corporate and government financial situations (i.e. Moody’s Corporation, Fitch Rating Ltd, and Standard & Poor’s a few among others).

3.2. Earlier techniques

As mentioned previously, the early studies concerning ratio analysis for bankruptcy prediction are known as the univariate studies. These studies consisted mostly of analyzing individual ratios, and sometimes, of comparing ratios of failed companies to those of successful firms. However, few studies were published up to the mid-60s4. This period is known as a relatively rich in published studies of corporate failures, in which academics advanced further in the field.

In particular, Beaver (1966) studied the predictive ability of accounting data as predictors of major events. His work was intended to be a benchmark for future investigations into alternative

---

4 See Horrigan (1968) and Bellovary et al. (2007) for further information in the early studies concerning corporate failure.
predictors of failure. Beaver found that a number of indicators could discriminate between matched samples of bankrupt and non-bankrupt firms for as long as five years prior to failure. In a real sense, his univariate analysis of a number of bankruptcy predictors set the stage for the development of multivariate analysis models.

Two years later, the first multivariate study was published by Altman (1968). With the well-known “Z-score”, which is a multiple discriminant analysis (MDA) model, Altman demonstrated the advantage of considering the entire profile of characteristics common to the relevant firms, as well as the interactions of these properties. Specifically, the usefulness of a multivariate model taking combinations of ratios that can be analyzed together in order to consider the context or the whole set of information at a time compared to univariate analysis that study variables one at a time and tries to gather most information at once. Consequently to this discriminatory technique, Altman was able to classify data into two distinguished groups: bankrupt and non-bankrupt firms. He also demonstrated a second advantage: if two groups were studied this analysis reduces the analyst’s space dimensionality to one dimension.

3.3. Evolution of statistical techniques

Altman’s works was then followed by subsequent studies that implemented comparable and complementary models. Meyer & Pifer (1970) employed a linear probability model (LPM). This is a special case of ordinary least square (OLS) regression with dichotomous (0-1) dependent variables for bank bankruptcy prediction. It is interesting to notice that while underlying assumptions of discriminant analysis and LPM are not similar, the results of the methods are identical.

Deakin (1972) compared Beaver’s and Altman’s methods using the same sample. He first replicated Beaver study’s using the same ratios that Beaver had used. Next, he searched for the linear combination of the 14 ratios used by Beaver which best predicts potential failure in each of five years prior to failure. Finally, he devised a decision rule, which was validated over a cross-sectional sample of firms. Deakin’s findings were in favor of the discriminant analysis, which compared to the univariate analysis, is a better classifier for potential bankrupt firms. The same year, Edmister (1972) tested a number of methods of analyzing financial ratios to predict small business failures. Even though he found that not all methods and ratios could be used as
predictors of failure, he confirmed that some ratios variables could be used to predict failure of small business companies. Finally, Edmister recommended using at least three consecutive year’s financial statement to predict small businesses bankruptcies.

Altman et al. (1977) constructed a new bankruptcy classification model called the “Zeta model” to update the “Z-score”. In particular, they compared linear and quadratic discriminant analyses for the original and holdout samples, introduced prior probabilities of group membership and costs of error estimates into the classification rule, as well as a comparison of the model’s results with naïve bankruptcy classification strategies. Altman et al. obtained good results with a classification accuracy: above 95% one period prior to bankruptcy and above 70% prior to five annual reporting periods.

Martin (1977) also presented a logistic regression model to predict probabilities of failure of banks based on the data obtained from the Federal Reserve System data sample. Martin was then followed by Ohlson (1980) who developed a logistic regression model, logit model or logit analysis (LA), to predict bankruptcies. He principally argued the MDA approach in regards of three following points: (i) the MDA technique relies too much on assumptions, (ii) the MDA outputs score do not provide intuitive interpretation, but however agreed that if a priori probabilities are known, then it becomes possible to derive a posteriori probabilities of bankruptcy, which is evident in the analysis sections of this study, and (iii) Ohlson pinpointed the discriminant selection process for its relative subjectivity. On the other hand, according to Ohlson, the use of conditional logit analysis avoids all of the problems discussed above. In particular, he underlines as major advantages that in logit model there is no need for assumptions to be made regarding a priori probabilities of bankruptcy, and for the distribution properties of the predictors. This approach model is in particular interesting because it allows the practitioner to test the significance of the predictors as it is presented in the assumptions’ test part of this study.

Zmijewski (1984) denounced that estimating models on nonrandom samples can result in biased parameter and probability estimates if appropriate estimation techniques are not used. Specifically, he presented two estimation biases: (i) one resulting from oversampling distressed firms, and (ii) the other from using only complete data. Zmijewski also used another interesting form of the logistic regression: the probit model or probit analysis (PA), to support his findings.
West (1985) used the combination of factor analysis (FA) and logit estimation as a new approach to measure the condition of individual institutions and to assign each of them a probability of being a problem bank. He demonstrated that the combination of factor analysis and logit estimation was promising in evaluating bank’s condition.

Karels & Prakash (1987) conducted a study in a threefold manner, (i) they investigated first if ratios used in previous firm failure studies satisfied the joint normality condition required by MDA technique, (ii) if these ratios were not normal, they constructed ratio sets that where multivariate normal or almost normal, and (iii) they then compared the newly built model with multivariate normal ratios to the results of other studies. Their results were not as expected. They could explain it because their model had too many different ratios to be comparable to others. Finally, because financial data are often non-normally distributed, Karels et al. underlined the fact that it would be better to use linear discriminant analysis (LDA) than quadratic discriminant analysis (QDA), which is too sensitive to the loss of the normality assumption.

Haslem et al. (1992) analyzed using a canonical analysis the foreign and domestic balance sheet strategies of the U.S. banks and their association to profitability performance based on a 1987 sample data. They found a consistent dichotomy in foreign and domestic asset and liability matching strategies, while domestic strategies appear more conservative with respect to interest-rate and liquidity risks. Banks that follow a predominant foreign strategy, compared to a domestic strategy, are found more profitable.

Altman (1993) adapted his “Z-score” to private firms’ application, which he called the “Z’-score”. This latest model differs from the original “Z-score” by substituting the book value of equity for the market value, and by re-estimating all the model’s coefficients.

Altman et al. (1995a) applied a further adaptation of the original “Z-score” to non-manufacturers and emerging markets’ firms, called the “Z’’-score” model. In this latest model, they decided to drop the asset turnover ratio in order to minimize the potential industry effect compared to the original “Z-score” model. Finally, they also re-estimated the model’s coefficients.

Few years later, Shumway (2001) developed a dynamic logit or hazard model for forecasting bankruptcy. Compared to the classic logit model that is based on single period data, the hazard model involves the modeling of multiple period data and in complement allows for time-varying
covariates. In addition, Shumway considered both classic accounting data and equity market data to form his model. In particular, he highlighted the usefulness of some previously neglected market driven variables such as: a firm’s market size, past stock returns, and the idiosyncratic standard deviation of firm stock returns, to forecast bankruptcy. He argued that his model is more consistent in predicting bankruptcy. Other recent studies using Shumway’s approach include Chava & Jarrow (2004), Hillegeist, Keating, Cram, & Lundstedt (2004), and Beaver, McNichols, & Rhie (2005).

Jones & Hensher (2004) developed a mixed logit model for financial distress prediction. Jones et al. argued that mixed logit model offers substantial improvements compare to binary logit and multinomial logit models (MNL). For example, in addition to fixed parameters, mixed logit models include estimates for the standard deviation of random parameters, the mean of random parameters, and the heterogeneity in the means as main improvements. They found that the out-of-sample accuracy of the mixed logit model was superior to the multinomial logit model.

Canbas et al. (2005) combined four different statistical techniques (PCA, DA, LA, and PA) to develop the integrated early warning system (IEWS) that can be used in prediction of bank failures. At first, principal component analysis (PCA) was used to explore the basic financial characteristics of the banks. Further on, discriminant analysis (DA), logit analysis (LA) and probit analysis (PA) models were estimated based on highlighted previous characteristics to construct the IEWS model. Results were in favor of the utilization of such a combination of four parametric approaches to the banking sector and more generally, they should be extended to other business sectors for failure prediction. The same year, Altman (2005) introduced the EMS model for emerging corporate bonds, which is an enhanced version of the “Z’’-score” model. This latest model as the advantage to be applicable to non-manufacturing companies and manufacturers, as well as being relevant for privately held and publicly owned firms.

Later on, Campbell, Hilscher, & Szilagyi (2008), implemented a dynamic logit model to predict corporate bankruptcies and failures at short and long horizons, using accounting and market variables. They argued empirical advantages of the model over the bankruptcy risk scores proposed by Altman (1968) and Ohlson (1980). Finally, they showed that stocks with a high risk of failure tend to deliver anomalously low average returns.
Recently, Altman, Fargher, & Kalotay (2011) estimated the likelihood of default inferred from equity prices, using accounting-based measures, firm characteristics and industry-level expectations of distress conditions. This approximately enables timely modeling of distress risk in the absence of equity prices or sufficient historical records of default. Model’s results are comparable to that of default likelihood inferred from equity prices using the Black-Scholes-Merton structure. Finally, Altman et al. emphasized the importance of treating equity-implied default probabilities and fundamental variables as complementary rather than competing sources of predictive information. In order to improve the analysis performance of logit model, Li, Lee, Zhou, & Sun (2011) presented a combined random subspace approach (RSB) with binary logit model (L) to generate a so called RSB-L model that takes into account different decision agents’ opinions as a matter to enhance results. Findings indicate that the newly proposed RSB-L model could be used as an alternative of classic statistical techniques in predicting corporate failure. J. Sun & Li (2011) tested the feasibility and effectiveness of dynamic modeling for financial distress prediction (FDP) based on the Fisher discriminant analysis model. They designed a framework of dynamic FDP based on various instance selection methods, such as full memory window, no memory window, window with fixed size, window with adaptable size, and batch selection. They also utilized initial features set composed of seven aspects of financial ratios and proposed a wrapper integrating forward and backward selection for the dynamic modeling of FDP. Findings indicated that dynamic models can perform better than static models and should be further developed to other classification techniques.

Finally, for additional readings on the subject of corporate bankruptcy related to this part, readers may refer to E. I. Altman & Edith Hotchkiss (2006), who present in a book several problematic related to the topic; E. I. Altman & Narayanan (1997), who present an international literature review of the topic; Beaver, Correia, & McNichols (2010), who in a monograph discuss the financial distress prediction literature, focusing on (i) the set of dependent and explanatory variables, (ii) the statistical methods of estimation, and (iii) the modeling of financial distress.

3.4. Alternative modeling techniques

In addition of statistical techniques, alternative modeling techniques to predict corporate bankruptcy have been largely developed and became popular in the recent years. In this part,
major techniques developed during the previous decades are presented such as: (i) neural networks, (ii) decision trees, (iii) case-based reasoning, (iv) operations research, (v) support vector machines, (vi) soft computing, and (vii) others.

3.4.1. Neural networks

Neural networks (NN) is probably the most widely used model among the intelligent techniques (Demyanyk & Hasan, 2010). Its principle is to mimic the biological neural networks of the human nervous system through an algorithm. This latest technique offers two interesting advantages compared to classic statistical techniques. The first one is that neural networks as non-parametrical models do not rely on specific assumptions like the distribution of predictors or properties of data. This makes it theoretically more reliable than models that would have their assumptions violated (as it is often the case and not the exception with financial data (Bardos (2001)). The other advantage is the reliance on nonlinear approaches, which offers extended possibilities for testing complex data patterns. A downside is that NN models may be more influenced by temporal or cyclical changes in the economy than classic statistical techniques (Bardos, 2001). Neural networks may also be difficult to interpret (Paliwal & Kumar, 2009). According to Ravi Kumar et al. (2007), the multi-layer perceptron (MLP), radial basis function network (RBFN), probabilistic neural network (PNN), cascade correlation neural network (Cascor), learning vector quantization (LVQ) and self-organizing feature map (SOM) are some of the popular neural networks architectures. These architectures differ mostly in their aspects such as the type of learning, node connection mechanism, or training algorithm for few examples. In the last two decades, many researchers have studied and developed neural networks models. For a complementary literature, readers may refer to Odom & Sharda (1990), E. Altman, Marco, & Varetto (1994), Wilson & Sharda (1994), Zhang (1999), Lee, Booth, & Alam (2005), and du Jardin (2010). In addition, Paliwal & Kumar (2009), reviewed articles that involve a comparative study of neural networks and statistical techniques used for predicting bankruptcy. In particular, Paliwal et al. presented their literature review according to different specific areas of research, such as (i) accounting and finance, (ii) medicine, (iii) engineering, (iv) marketing and (v) general applications.
3.4.2. Decision trees

Decision trees (DT) produce a set of if-then rules that divide a large heterogeneous data set into smaller, more homogenous groups with respect to a particular value of the target variable. Different algorithms can be used for building decision trees, such as classification and regression trees (CART), chi squared automatic interaction detection (CHAID), Quest, C4.5, C5.0, or entropy reduction algorithm (Ravi Kumar & Ravi, 2007). Decision trees have been popularly used for classification problems, because their rules are easy to understand and communicate (Cho, Hong, & Ha, 2010). However, they may not be as robust to cyclical changes as classic LDA (Bardos & Rasson, 2001). Several studies and researches have been conducted on this topic. For further literature, readers may refer to Marais, Patell, & Wolfson (1984), Frydman, Altman, & Duen-Li (1985), and Li, Sun, & Wu (2010).

3.4.3. Case-based reasoning approach

Case-based reasoning (CBR) can be explained as a similar process to the decision making process of the human being. The basic idea involves solving new problems based on previous cases and their solutions. The solution algorithm of CBR approach is based on a distance function and on a combination function. The distance function (i.e. Euclidean distance) calculates the distance between two records, and the combination function combines the results from several neighbors (i.e. k nearest neighbor) to arrive to an answer. An interesting feature of this technique is that solutions are very comprehensive and can be reused directly or indirectly to possibly solve newly encountered problems (Li & Sun, 2008). This technique was firstly introduced into the domain of business failure prediction by researches like Jo & Han (1996), Jo, Han, & Lee (1997), and Bryant (1997). Results from their studies did not provide enough evidence that CBR models were more applicable than other reference models. However, some researchers have kept and demonstrated an interest in this technique, attempting to improve its initial predictive performance. For further literature, readers may refer to Park & Han (2002), Yip (2004), Li & Sun (2009), and Li & Sun (2011c).

3.4.4. Operations research

Originating in military efforts before World War II (Gass & Assad, 2005), operations research is an interdisciplinary mathematical science that focuses on the effective use of technology by organizations. Operations research applies mathematical programming techniques to decision
making, aiming at optimal or near-optimal solutions to complex problems. Mathematical programming (MP) techniques compare to statistical methods offer three main advantages (M. Sun, 2011). First, as nonparametric methods, MP techniques are not relying on strict assumptions such as statistical techniques do. Further, MP techniques are also able to perform correctly on a broader variety of data. Finally, the fitted model in MP techniques is less influenced by any outlier observations. Different techniques and models have been introduced in the literature. One of the first to introduce linear programming approaches to the classification problem were Freed & Glover (1981a). Their work was then followed by subsequent studies that implemented comparable and complementary models, such as: linear programming (LP) (Freed & Glover, 1986b; Kwak, Shi, & Kou, 2011); nonlinear programming (A. Stam & Joachimsthaler, 1989); linear goal programming (LGP) (Freed & Glover, 1981b, Gupta, P. Rao, & Bagchi, 1990); integer programming (IP) (Glen, 1999); mixed integer programming (MIP) (Xu & Papageorgiou, 2009); data envelopment analysis (DEA) (Cileen, 2004) among others. In general, findings show that mathematical programming approaches can perform as good as traditional statistical techniques (Kwak et al., 2011). In particular, MP approaches may be preferred when assumptions underlying the statistical approaches are seriously violated (A. Stam, 1990) or (Ragsdale & Stam, 1991). However, (M. Sun, 2011) pointed out that researchers and practitioners will be more willing to accept MP approaches as nonparametric procedures when simple but powerful multiple-class MP models will be available.

3.4.5. Support vector machines

Support vector machine (SVM) is one of the latest techniques developed and implemented to predict corporate bankruptcy. Introduced by Boser, Guyon, & Vapnik, (1992) and Vapnik & Cortes (1995), the basic idea of SVM is to map the input vector into some high dimensional feature space through some nonlinear mapping chosen a priori. In this space a linear decision surface is constructed with special properties that ensure high generalization ability to network. SVM is gaining popularity due to many attractive features and excellent generalization performance on a wide range of problems. In particular, this technique offers two major advantages: (i) it takes linear non-separable situations into account, which extends the model’s possibilities and flexibility in finding suitable or undiscovered variables in predicting bankruptcy, and (ii) it adopts the principle of structural risk minimization that reduces over fitting the model on the training data set for a stronger classifying ability (S. Chen, Härdle, & Moro, 2011).
However, one of the principle drawbacks of this method is that it procures little explanation on variables contributing to a bankrupt (Kaya, Gurgen, & Okay, 2008). Therefore, this method may offer superior predictive abilities but may not be preferred by practitioners attempting to fix a potential bankruptcy (at least in a simple stand-alone mode). Some literature on the topic can be found at: Min & Lee (2005), Hua, Wang, Xu, Zhang, & Liang (2007), Trustorff, Konrad, & Leker (2010), and Li & Sun (2011d).

3.4.6. Soft computing

Soft computing\(^5\) combines several individual techniques to maximize their advantages while it minimizes combined model’s weaknesses. The general idea is that the gains achieved by precision and certainty, as in more conventional methods (i.e. LDA, Logit, NN, etc.), are not justified by their costs (Ravi Kumar & Ravi, 2007). This technique has recently become very popular among researchers and practitioners and is seen as one of the latest trend in corporate prediction modeling (Demyanyk & Hasan, 2010). There are many different possibilities of combinations and associations. Combinations of techniques are not exclusively reserved to solely artificial intelligent techniques, which are often found complementary (Ravi Kumar & Ravi, 2007). Statistical techniques, operations research, as well as other techniques found useful in predicting bankruptcies can be combined to develop the ultimate model. For instance, combinations of statistical techniques are frequently accompanied by artificial intelligence systems for better model performance in practice. For example, Huang, Tsai, Yen, & Cheng (2008) present a hybrid financial analysis model including static and trend analysis models to construct and train back-propagation neural network (BPN) model. Their results outperform other models including discriminant analysis, decision trees, and the back-propagation neural network model alone. Other developed models include: hybrid case-based reasoning and genetic algorithm (Ahn & Kim, 2009), combined six different classification algorithms such as MDA, Logit, NN, DT, SVM, and CBR (J. Sun, Li, & Zhang, 2009), principal component analysis with multivariate discriminant analysis and logistic regression (Li & Sun, 2011a), principal component case-based reasoning ensemble (Li & Sun, 2011e). Finally, for a survey, readers may refer to Verikas, Kalsyte, Bacauskiene, & Gelzinis (2010).

---

\(^5\) One of the first to coin this term was Zadeh (1965)
3.4.7. Other techniques

Several models were quoted in this study. However, the presented techniques are not exhaustive\(^6\) and there are many different models not covered in this review that have been implemented to test bankrupt prediction. Some of these models are: genetic algorithm (GA) (Varetto, 1998) (Davalos, Gritta, & Adrangi, 2010), fuzzy set theory (Zadeh, 1965) (Zarei, Rabiee, & Zanganeh, 2011), rough sets (Pawlak, 1982) (Mosqueda, 2010), gaussian processes (Peña, Martínez, & Abudu, 2011), isotonic separation (Ryu & Yue, 2005), gambler’s ruin Model (Wilcox, 1971), option pricing theory (Merton, 1974), cash flow models (Gentry, Newbold, & Whitford, 1987) (Aziz, Emanuel, & Lawson, 1988), return variation models (E. I. Altman & Brenner, 1981) (Clark & Weinstein, 1983), risk index models (Tamari, 1966) (Moses & Liao, 1987), AdaBoost (J. Sun, Jia, & Li, 2011) (Moses & Liao, 1987), trait recognition (Kolari, Caputo, & Wagner, 1996), self-organizing learning array (SOLAR) method (Zhu, He, Starzyk, & Tseng, 2007), dynamic modeling techniques (J. Sun, He, & Li, 2011) (J. Sun & Li, 2011; J. Sun et al., 2011; J. Sun, Jia et al., 2011). Finally, for further information on the complex and wide literature on alternative modeling techniques of bankruptcy prediction, readers may refer to Ravi Kumar & Ravi (2007), who present a comprehensive review of the researches done from 1968 to 2005 in the application of statistical and intelligent techniques applied to solve the bankruptcy prediction problem faced by companies; to Demyanyk & Hasan (2010), who provide a summary of the empirical results obtained in several economics and operations research papers that attempt to explain, predict, or suggest remedies for financial crises or banking defaults; and to S. Balcaen & Ooghe (2004) for further literature review on the topic.

3.5. Evolution and empirical applications in France

In France, the research was initiated on a cooperative basis by a team of researchers (E. I. Altman, Margaine, Schlosser, & Vernimmen, 1974) with the assistance of the French Central Bank\(^7\). Altman et al. developed a model for determining the credit worthiness of commercial loan applicants applied to the French textile industry, which was suffering major competition by that time. Altman et al. assessed the combined potential of traditional financial statement analysis

---

\(^6\) According to Bellovary et al. (2007) study, there are over 150 models available to predict corporate bankruptcy and many of them have demonstrated high predictive abilities.

\(^7\) Banque de France
with several relatively modern statistical procedures. In particular, they investigated the global nature of a large number of financial ratios through the use of a principle component analysis (PCA). Further, the most important financial indicators detected were processed through a multiple linear discriminant model to assess the credit worthiness of commercial loans applicants. Results were not as high as expected and the model was not implemented on a practical basis (E. I. Altman & Narayanan, 1997). However, most financial ratios seemed to discriminate well between good and bad credit risks based on their mean values, providing interesting insights of that particular troubled industrial sector.

Following Altman et al. initial work, researches and studies in France were mostly conducted on behalf of the French Central Bank as well as by some independent researchers such as: Mader (1975), Collongues (1977), and Conan & Holder (1979), to quote a few among others.

In 1982 the first operational score usable by practitioners such as banks or companies was developed by the French Central Bank for the industrial sector (Bardos, 1998a). The same year, Zollinger (1982) analyzed the risk of corporate credit based on the Electre method: a multi-criteria outranking method. Introduced by B. Roy, Benayoun, & Sussman, (1966) and B. Roy (1968), this approach is also known as the French school of decision making (Dimitras, 1996).

In the following years, scores were applied and ameliorated to other sectors. Micha (1984) presented three main objectives that were set to face business failure in France. Researchers would have: (i) to find a robust discriminant function capable of discriminating firms up to three years in advance, (ii) to have a model that can be applied for prediction, and (iii) to formulate a model that is exclusively based on quantitative data (i.e. accounting, economic and financial data) to guarantee a maximum of objectivity to calculate and analyze the function.

Bardos (1989) compared the linear discriminant analysis, the logistic regression, CART, and the disqual methodology. The latest technique was introduced by Saporta (1977). It is a discriminatory technique based on qualitative variables. As regard to the results, the linear discriminant analysis was chosen for its robustness to cyclical temporal changes, as well as for its good interpretability and maintenance.

Bardos & Zhu (1997) analyzed and compared results from three discriminatory techniques, such as a Fisher linear discriminant analysis, a logistic regression and a multilayer neural network
method. To facilitate the method comparison, data utilized were the same in all three methods. Bardos et al. concluded that neural networks produce similar results as linear discriminant analysis and logistic regression. However, the linear discriminant analysis method appeared to be the best at matching temporal stability, which in the studies’ condition made the LDA method the most robust technique to cyclical and economical changes.

During the last decade, studies on industry focused model were developed such as: Bardos (1998b) who developed a model for industries, Stili (2002) who developed a model for the construction industry, and Planès (2004) who developed a model for the hotel and restaurant industry. Bardos (2005) reviewed all the scores available at the French Central Bank. She presented the latest score applications, researches that the Central Bank offers, as well as the latest updates regarding the sectorial default rate and the probability of default to a three year horizon, according to each class of risk. This review is often updated and from the latest version to this date, the “Banque de France” has developed eight categories of scores to predict bankruptcies, which are described in the next part that is specific to focused models.

Finally, for further literature on the specific evolution in France, readers may refer to Bardos (2001), who summarized in a comprehensive book most of the fundamentals that have been implemented so far in France regarding prediction of corporate bankruptcy. In particular, she devoted whole chapters to linear discriminant analysis, to logistic regression, and to decision trees, as well as a brief explanation of other techniques such as neural networks and the disqual methods. In her book, she also dedicated a chapter to the selection of the variables and database and the final validation of the model. In addition, readers may refer to Refait-Alexandre (2004), who presented an overall literature review on corporate bankruptcy from a French perspective. She is also confirming the usefulness of the LDA technique in predicting corporate bankruptcy, in particular from an operational point of view.

3.6. Focused models

Most of the models in the literature are general models developed for multiple industries (often medium to large size companies). On the other hand, focused models are specific to an industry and firm size. Compared to a general model, focused models do not appear to follow a special trend and are rather developed on academics’ need (Bellovary et al., 2007). For example,
academics have developed focused models for SMEs (E. I. Altman & Sabato, 2007), hotel and lodging industry (Youn & Gu, 2010) (Li & Sun, 2011b) (Kim, 2011), internet firms (Chandra, Ravi, & Bose, 2009) (Ravisankar, Ravi, & Bose, 2010), construction (J. Chen, 2011), or family owned businesses (Konstantaras & Siriopoulos, 2011), among others.

Researchers believe that with a focused model they will increase and obtain better results than with a general model. Altman et al. explained the advantages of an industry focused model as:

“...models developed for specific industries (e.g., retailers, telecoms, airlines, etc.) are even better method for assessing distress potential of like-industry firms.” (E. I. Altman & Edith Hotchkiss, 2006), page 249

Agreeing with Altman quote, several industry specific scores have been constructed by the French Central Bank, which are in their chronological order of appearance: (i) BDFI2 that is for industrial companies (since 2003), (ii) BDFT2 that is for transportation industry (since 2003), (iii) BDFCG that is for the wholesale industry (since 2003), (iv) BDFCD that is for the retail business and auto repair industry (since 2003), (v) BDFH2 that is for the lodging and hotel industry (since 2005), (vi) BDFR2 that is for the restaurant industry (since 2005), (vii) BDFSA/B that are for business services industries (since 2005), and (viii) BDFB2 that is for the construction industry (since 2009).

In this study is developed a focused model for SMEs in the French hospitality/ accommodation industry (NAF: 55). For this matter, a linear discriminant analysis method is used. This method appears to be one of the most appropriate techniques in predicting corporate bankruptcy for the study’s specific. The theoretical methodology of the linear discriminant analysis is presented in the section that follows.

---

8 For further information, see « Les scores de la Banque de France : leur développement, leurs applications, leur maintenance », updated version October 2010.
4. **Theoretical model**

The linear discriminant analysis (LDA) is a statistical technique used to separate (discriminate) groups from a population. This technique was originally introduced in the biological science by Fisher (1936), who distinguished three species of Iris flowers based on group characteristics.

The LDA is one of the most useful techniques to discriminate and predict corporate bankruptcies, in particular when there are solely quantitative predictors. Compared to other techniques such as logistic regression, classification trees, neural networks, and others, LDA has the advantages to be robust against a certain degree in the loss of assumptions, to relatively resist to temporal changes and to offer judicious possibilities of interpretation.

Once data are collected, the statistical analysis is composed of two successive steps: descriptive and inferential. The descriptive step determines the separation representation between the preexisting \( g \) groups based on the training data. The inferential or decisional step consists in elaborating the decision rule to classify new objects (firms).

The LDA can be implemented according to two decision rule approaches: (i) geometrical and (ii) probabilistic.

4.1. **Geometrical approach**

The geometrical approach relies on a metrical rule to separate at best preexisting groups (bankrupt and non-bankrupt firms) in a Cartesian coordinate system. The separation points will bring closer the representative points of the objects of the same group and set apart the representative points of the objects from different groups. Therefore, the separator hyperplane maximizes inter-groups variances (between) and minimizes intra-groups variances (within).

In the case of two groups \( i \) and \( j \) (bankrupt and non-bankrupt firms), the optimal separator hyperplane under the metric criterion has for equation:

\[
(\mu_i - \mu_j)'M\left(x - \frac{\mu_i + \mu_j}{2}\right) = 0, \forall i \neq j
\]  

(4.1)
where \( \mu_i \) and \( \mu_j \) are the means of groups \( i \) and \( j \). \( M \) is a metric used to measure the relative distance. \( M \) is often used as the inverse total covariance matrix or as the inverse intra-class covariance matrix. \( x \) is the vector of the \( k \) ratios of the firm.

This explanation involves a notion of distance to assess the relative proximity and distance of the point cloud. This distance is essential in classifying new objects. For example, if a new object (firm) for which the descriptive variables as financial ratios are known but the classification group is unknown, the geometrical rule of classification affects this object into the group whose average point is the closest to the representative point of the object. The new object characterized by \( x \) is affected into the group \( i \) if and only if the distance \( d(x, \mu_i) \) is strictly inferior to the distance \( d(x, \mu_j) \), such as:

\[
d(x, \mu_i) < d(x, \mu_j), \forall i, j \in \{1, 2, ..., g\} \text{ and } i \neq j \tag{4.2}
\]

For example, Figure 1 illustrates the theory of linear discriminant analysis presented above for the classification of two groups with two descriptive variables \( (x_1 \text{ and } x_2) \). In the Cartesian coordinate system, all variables or individual multivariate characteristics \( (x_1 \text{ and } x_2) \) plotted in a \( k \) dimensional space (2 dimensions here), are transformed by the discriminant function \( f(z) \) – Equation (4.1) into a single one dimensional output with the z score (located along the line z). Applying the classification rule as depicted in Equation (4.2), if \( D(z) < 0 \) – the object is affected to group \( i \); if \( D(z) > 0 \) – the object is affected to group \( j \); for \( D(z) = d(x, \mu_i) - d(x, \mu_j) \).

However, this approach, purely geometrical, does not consider the a priori probabilities of the different groups and their potential cost of misclassification, while the probabilistic classification approach offers such possibilities.
4.2. Probabilistic approach

In probabilistic models, each observation $x$ of the training data is no longer considered as a Cartesian coordinate but as the realization of the object’s description. Each different group of object is considered by its a priori probability of appearance, which limits the possibility of appearance of different objects to be classified.

Knowing the group membership and description of objects $x$, it is possible to estimate the probability that a particular description is realized as the group $i$ of the object, such as: $P(x|i)$ with $i \in \{1,2,\ldots,g\}$. Therefore, it is assumed that certain descriptions have more chances to be
realized for some groups than others based on their distributional differentiation. It is equivalently assumed that both groups have proper characteristics and that each object (bankrupt or non-bankrupt firm) presenting these characteristics is affected in the same classification group.

The decision criterion delimits the separation between the studied groups. Therefore, with the probability \( P(x|i) \) it becomes possible to classify objects according to their descriptions \( x \) in the group for which the probability that this description is achieved is maximum; leading to the rule that the object of description \( x \) is affected to group \( i \) if and only if the probability \( P(x|i) \) is strictly superior than the probability \( P(x|j) \) for all \( j \in \{1, 2, ..., g\} \) and with \( i \neq j \), such as:

\[
P(x|i) > P(x|j) \quad \forall \; i, j \in \{1, 2, ..., g\} \quad \text{and} \quad i \neq j
\]  

(4.3)

However, it would be preferable to obtain the probability of belonging to a group experiencing the description of interest, leading to the rule: \( P(i \mid x) > P(j \mid x) \) \( \forall \; i, j \in \{1, 2, ..., g\} \) and \( i \neq j \), rather than that a particular description is realized in a given group such as in Equation (4.3).

Bayes’ theorem allows this interchange, such as:

\[
\begin{align*}
P(i \mid x) &= \frac{P(x \mid i).P(i)}{\sum_{n=1}^{g} P(x \mid n).P(n)} \\
P(j \mid x) &= \frac{P(x \mid j).P(j)}{\sum_{n=1}^{g} P(x \mid n).P(n)}
\end{align*}
\]

Therefore, Equation (4.3) can be rewritten such as:

\[
\frac{P(x \mid i).P(i)}{\sum_{n=1}^{g} P(x \mid n).P(n)} > \frac{P(x \mid j).P(j)}{\sum_{n=1}^{g} P(x \mid n).P(n)}
\]

which produces

\[
P(x \mid i).P(i) > P(x \mid j).P(j), \forall \; i \neq j
\]  

(4.4)

Consequently, the classification rule consists of maximizing the probability that an object belongs to a group according to its descriptions.
4.2.1. Assumption of multivariate normality

In practice and when the descriptors are continuous, it is often assumed that the descriptions of each group follow a normal distribution. The distributional structure of descriptors is therefore differentiated by the parameters of the law, while variables $x$ of group $i$ are assumed to follow the normal law, such as:

$$P(x|i) = \left(\frac{1}{(2\pi)^{\frac{k}{2}}|W_i|^\frac{1}{2}}\right) \exp \left[-\frac{1}{2}(x - \mu_i)'W_i^{-1}(x - \mu_i)\right]$$

where $W_i$ is the covariance matrix of group $i$ and $\mu_i$ the mean vector of group $i$.

Therefore, Equation (4.4) can be rewritten as:

$$\left(\frac{P(i)}{(2\pi)^{\frac{k}{2}}|W_i|^\frac{1}{2}}\right) \exp \left[-\frac{1}{2}(x - \mu_i)'W_i^{-1}(x - \mu_i)\right] > \left(\frac{P(j)}{(2\pi)^{\frac{k}{2}}|W_j|^\frac{1}{2}}\right) \exp \left[-\frac{1}{2}(x - \mu_j)'W_j^{-1}(x - \mu_j)\right]$$

which becomes after subtracting the $(2\pi)^{\frac{k}{2}}$ on each side

$$\left(\frac{P(i)}{|W_i|^\frac{1}{2}}\right) \exp \left[-\frac{1}{2}(x - \mu_i)'W_i^{-1}(x - \mu_i)\right] > \left(\frac{P(j)}{|W_j|^\frac{1}{2}}\right) \exp \left[-\frac{1}{2}(x - \mu_j)'W_j^{-1}(x - \mu_j)\right]$$

Then, taken to the logarithm this inequation produces

$$\ln(P(i)) - \frac{1}{2}\ln(|W_i|) - \frac{1}{2}(x - \mu_i)'W_i^{-1}(x - \mu_i)$$

$$> \ln(P(j)) - \frac{1}{2}\ln(|W_j|) - \frac{1}{2}(x - \mu_j)'W_j^{-1}(x - \mu_j)$$

which after multiplying both side by 2 becomes

$$\leftrightarrow -2\ln(P(i)) + \ln(|W_i|) + (x - \mu_i)'W_i^{-1}(x - \mu_i)$$

$$< -2\ln(P(j)) + \ln(|W_j|) + (x - \mu_j)'W_j^{-1}(x - \mu_j)$$

(4.5)

Consequently, without further assumptions than the normality of predictors, the assignment rule of minimum risk is quadratic and the border between the regions allocation is also quadratic.

Thus, if

$$d_i(x) = \ln(|W_i|) + (x - \mu_i)'W_i^{-1}(x - \mu_i)$$

$$d_j(x) = \ln(|W_j|) + (x - \mu_j)'W_j^{-1}(x - \mu_j)$$

the discriminant function is quadratic.
Thus, Equation (4.5) can be rewritten as:

\[
d_i(x) - 2\ln(P(i)) < d_j(x) - 2\ln(P(j))
\]

4.2.2. Assumption of homoscedasticity

In addition, if all covariance matrices are assumed to be equal for all groups, such as \( W_i = W_j = W \), the assignment rule becomes linear.

Therefore, Equation (4.5) can be rewritten as:

\[
-2\ln(P(i)) + \ln(|W|) + (x - \mu_i)'W^{-1}(x - \mu_i) < -2\ln(P(j)) + \ln(|W|) + (x - \mu_j)'W^{-1}(x - \mu_j)
\]

on both side of the inequation \( \ln(|W|) \) can be subtracted

\[
-2\ln(P(i)) + (x - \mu_i)'W^{-1}(x - \mu_i) < -2\ln(P(j)) + (x - \mu_j)'W^{-1}(x - \mu_j)
\]

developing this inequation produces

\[
-2\ln(P(i)) - 2\mu_i W^{-1}x' + \mu_i W^{-1} \mu_i > -2\ln(P(j)) - 2\mu_j W^{-1}x' + \mu_j W^{-1} \mu_j
\]

after multiplying both side by \(-\frac{1}{2}\) produces the linear discriminant function as follows

\[
\ln(P(i)) + \mu_i W^{-1}x' - \frac{1}{2} \mu_i W^{-1} \mu_i > \ln(P(j)) + \mu_j W^{-1}x' - \frac{1}{2} \mu_j W^{-1} \mu_j \quad (4.6)
\]

4.2.3. Assumption of equality of a priori probabilities of bankruptcy (and misclassification costs)

Finally, if all a priori probabilities are assumed to be equal, such as \( P(i) = P(j) \), Equation (4.6) becomes:

\[
\mu_i W^{-1}x' - \frac{1}{2} \mu_i W^{-1} \mu_i > \mu_j W^{-1}x' - \frac{1}{2} \mu_j W^{-1} \mu_j \quad (4.7)
\]

And the linear discriminant functions become:

\[
\begin{aligned}
f_i &= \mu_i W^{-1}x' - \frac{1}{2} \mu_i W^{-1} \mu_i', \\
f_j &= \mu_j W^{-1}x' - \frac{1}{2} \mu_j W^{-1} \mu_j'
\end{aligned}
\]
Therefore, the object characterized by $x$ is affected in the group $i$ if and only if the function $f_i$ is strictly superior to the function $f_j$ such as:

$$f_i > f_j, \forall \ i \neq j$$  \hspace{1cm} (4.8)

4.2.4. First formulation of the score function: the grouped classification

Classification of objects in group $i$ and $j$ is now possible. However, it would be more interesting to define a score function that regroups $f_i$ and $f_j$, such as Equation (4.9).

Hence, Equations (4.7) and (4.8) can be rewritten in a form that provides directly a classification such as:

$$\left((\mu_i - \mu_j)W^{-1}.x - (\mu_i - \mu_j)W^{-1}\left(\frac{\mu_i + \mu_j}{2}\right)\right) > 0, \forall \ i \neq j$$  \hspace{1cm} (4.9)

Equation (4.9) can be generalized in the following score function:

$$f(x) = \alpha'x + \beta$$  \hspace{1cm} (4.10)

where $\alpha' = (\mu_i - \mu_j)W^{-1}$ is the $k$ coefficient vector, $\mu_i$ and $\mu_j$ are the means on each group, $W^{-1}$ is the intra-class covariance matrix, $x$ is the vector of the $k$ ratios of the firm, and $\beta = -(\mu_i - \mu_j)'W^{-1}\left(\frac{\mu_i + \mu_j}{2}\right)$ is a constant.

The score function can be further developed in the following practical form:

$$f(x) = \alpha_1x_1 + \alpha_2x_2 + \cdots + \alpha_kx_k + \beta$$

where $\alpha_k = (\alpha_1,\alpha_2,\cdots,\alpha_k)$ is the $k$ coefficients vector, $x = (x_1,x_2,\cdots,x_k)$ is the vector of the $k$ ratios of the firm, and $\beta$ is a constant.

4.2.5. Second formulation of the score function: the variables contributions

The score function presented in the previous Equation (4.10) solely classifies objects in different groups. It would be more interesting to understand what the reasons of a certain classification and score are. In particular, the interpretability of the model would be enhanced if that would be possible to distinguish variables that raise the score to the one that lower it. Fortunately, one of
the advantages to use the linear discriminant analysis technique is that it makes the interpretability of the score easier, such as Equation (4.9) can be rewritten as:

\[(\mu_i - \mu_j)’W^{-1}\left(x - \frac{\mu_i + \mu_j}{2}\right) > 0, \forall \ i \neq j\] (4.11)

It is interesting to notice that this Equation (4.11) is the same as Equation (4.1), which has its assignment rule based on distances, as mentioned previously.

The general score function can be obtained from equation (4.11):

\[f(x) = \alpha’(x - P)\] (4.12)

where \(\alpha’ = (\mu_i - \mu_j)'W^{-1}\) is the k coefficient vector, \(\mu_i\) and \(\mu_j\) are the means on each group, \(W^{-1}\) is the intra-class covariance matrix, \(x\) is the vector of the \(k\) ratios of the firm, and \(P^k = \left(\frac{\mu_{i}^k+\mu_{j}^k}{2}\right)\) is the pivot value for the ratio \(k\).

The Decision rule is: if \(f(x) > 0\) then the object (firm) is allocated to the group \(i\), otherwise the object is allocated to group \(j\).

Equation (4.12) in its developed form can be written as:

\[f(x) = \alpha_1(x_1 - P_1) + \alpha_2(x_1 - P_1) + \cdots + \alpha_k(x_k - P_k)\]

where \(\alpha_k = (\mu_{i}^k - \mu_{j}^k)'W^{-1}\) is the \(k\) coefficients vector, \(x = (x_1, x_2, \ldots, x_k)\) is the vector of the \(k\) ratios of the firm, and \(P^k = \left(\frac{\mu_{i}^k+\mu_{j}^k}{2}\right)\) is the pivot value for the ratio \(k\).

Therefore the expression between the brackets multiplied to the associated coefficient, such as: \(C_k = \alpha_k(x_k - P_k)\) represents the contribution of \(k^{th}\) descriptor to the score’s value of \(f(x)\), conditionally to other variables.

Finally, the different contributions indicate how good the firm is performing on each studied variables. For example, a contribution that is negative highlights a weakness in a firm specific aspect (e.g. profitability, liquidity, etc.); rather than a contribution that is positive refers to a stronger characteristic.

The next section of this study concerns the data section.
5. Data

Doing linear discriminant analysis (LDA) is a difficult and long exercise. Every practitioner agrees that the data preparation is quite exhaustive and counts as a big part in the analysis. However, it is highly important to carefully do the data part. Because if a better data gets in the model, the final outcome will be better. For example, Altman said:

“... there is nothing more important in attracting rigorous and thoughtful research than data!” ((Altman & Edith Hotchkiss, 2006), page vii)

Therefore, 74 ratios are carefully preselected to build the best model. All these ratios cover different aspects of a company’s business such as: (i) liquidity, (ii) leverage, (iii) solvability, (iv) profitability, (v) operating performance measurement, (vi) efficiency, (vii) cash flow ratios, and (viii) asset utilization measurements. Appendix A presents the detailed list of ratios and their descriptions. These ratios are determined according to their popularity in the literature as well as to their subjectively judged potential. Among others, in particular the following sources are quite useful to determine and select ratios for the study: Soffer et al. (2003), Bragg (2006), Hansen & Mowen (2007), Ravi et al. (2007), Brealey, Myers, & Marcus (2009), Ross, Westerfield, & Jordan (2010), Porter & Norton (2011), and Brealey, Myers, & Allen (2011).

In general, the ratios that measure and distinguish the most significantly bankrupt to non-bankrupt firms are ratios measuring liquidity, solvency and profitability. However, their order of importance differs considerably since each study depicts a different association of ratios as being the most effective.

This section is organized into the following two parts: first, are presented the database and its particularities such as: (i) the database used, (ii) the definition of bankruptcy and firm’s status, (iii) other firm’s specificities and limitations of the database, and (iv) descriptive statistics. Then, reprocessing of the data and its different techniques are presented.
5.1. Data specificities

5.1.1. Database

In statistic, a situation is first modeled based on data that represent a whole population. This representativeness is key in the development and performance of the developed tool, study and model. It is important to conserve the characteristics of the population studied, on both training and testing data developed.

Using the global companies database Orbis from Bureau van Dijk (BvD), selection of firms that are bankrupt and non-bankrupt is possible. The ability to distinguish bankrupt to non-bankrupt firms is a fundamental factor.

5.1.2. Definition of bankruptcy and firm’s status

Another fundamental factor is to define the term – Bankruptcy:

Bankruptcy term as used in this study is defined by the fact that a firm should have ceased any activity because it is in the impossibility to pay its due debts to its creditors.

The Orbis database from BvD offers several options to limit and select the “firm’s status”. For bankrupt firms, for example, it is possible to select different status. In accordance of the definition of bankruptcy previously mentioned, firms that are defined as “bankrupt”, according to the BvD option scale, are selected. The other options are not considered in this study.

In addition, only non-bankrupt firms marked as “active” in the Orbis’ classification are considered. Classified “Active” companies compared to the other classifications (see Appendix C), represents the healthiest firms in the database. Therefore, it should be helpful to select the healthiest firms among non-bankrupt firms in order to ease the comparison versus bankrupt firms.

In summarize, only firms that are “Active” or “Bankrupt” in their status are considered in this study. Further descriptions of the BvD’s definition of the firm’s status are presented in the Appendix C.

5.1.3. Firm’s specificities and data limitations

This study focuses on the bankruptcy of the SMEs. Larger and multinational companies are not considered, because big companies (or multinationals) are often present in different countries
with different activities. Therefore, it is difficult to compare them with small or local competitors that have restricted possibilities. Following are presented seven limitations that are applied in this study such as: (i) industry’s specificity, (ii) period studied, (iii) operating revenues, (iv) assets restrictions, (v) location, (vi) number of employees, and (vii) ownership’s profile.

**Industry selected and NAF Rev. 2 code**

In November 2009, the French Central Bank counted eight “Score Banque de France”. These scores cover the following industries: (1) Manufacturing, (2) Retail and Car Repair, (3) Wholesale, (4) Transportation, (5) Construction, (6) Hospitality/Accommodation, (7) Cafes and Restaurants, (8a) Business Services – subsector A, and (8b) Business Services – subsector B.

All these sectors have similarities and differences. This study focuses on companies operating in the hospitality/ accommodation industry (No. 6 as mentioned previously). There are mainly two reasons for this choice. The first one is that the industry selected provides sufficient amount of data to perform the analysis. The other reason is that the model focuses on this industry in order to give better results.

Precisely, firms that are considered in this study are part of the NAF 55 code. The NAF is the French system for classifying industries. It is the equivalent of the SIC in the United States for example. According to the French National Institute for Statistics and Economics Studies (INSEE), the Nomenclature NAF 55 is organised as follow:

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>Accommodation</td>
</tr>
<tr>
<td>55.1</td>
<td>Hotels and similar accommodation</td>
</tr>
<tr>
<td>55.10</td>
<td>Hotels and similar accommodation</td>
</tr>
<tr>
<td>55.10Z</td>
<td>Hotels and similar accommodation</td>
</tr>
<tr>
<td>55.2</td>
<td>Holiday and other short-stay accommodation</td>
</tr>
<tr>
<td>55.20</td>
<td>Holiday and other short-stay accommodation</td>
</tr>
<tr>
<td>55.20Z</td>
<td>Holiday and other short-stay accommodation</td>
</tr>
<tr>
<td>55.3</td>
<td>Camping grounds, recreational vehicle parks and trailer parks</td>
</tr>
<tr>
<td>55.30</td>
<td>Camping grounds, recreational vehicle parks and trailer parks</td>
</tr>
<tr>
<td>55.30Z</td>
<td>Camping grounds, recreational vehicle parks and trailer parks</td>
</tr>
<tr>
<td>55.9</td>
<td>Other accommodation</td>
</tr>
<tr>
<td>55.90</td>
<td>Other accommodation</td>
</tr>
<tr>
<td>55.90Z</td>
<td>Other accommodation</td>
</tr>
</tbody>
</table>

Source: INSEE

As mentioned in the literature review, focused models may lead to better results.
**Period studied**
To insure temporal robustness to economic changes, it is recommended to build and study the model on a relatively long period, more exposed to potential changes in conjuncture (Bardos & Rasson, 2001). Therefore, in this study is considered the whole decade from year 2001 to 2010.

**Operating revenues**
Firms having their operating revenues included in the range that follows are considered: the minimum is 100,000 Euros and the maximum is 10,000,000 Euros.

**Assets restrictions**
This study is limited to firms having their assets bellow 20,000,000 Euros.

**Location selected for the study: French national market**
The world is vast and complex. Companies do not react and operate the same in different geographic areas. This work focuses on companies operating at national level. A market to make this study with satisfactory amount of data and with some understanding is the French market.

**Number of employees**
This study is not limited by the number of employees because this type of data is not available for all firms. However, the most of the firms have comparable number of employees (between 10 to 50 employees).

**Ownership’s profile**
Exclusively for the active firms, the study limits the ultimate owner. This is because the most of the active firms, even if small, are owned by a bigger firm and can be considered as subsidiaries of a multinational. To make the comparison more equal between bankrupt and active\(^ {10} \) firms it is better to restrict the number of active firms to the ones that are held by small investors, families or independent competitors. In detail, selected companies are owned by an ultimate owner that possesses at least 50.01% of the shares of the company.

### 5.1.4. Data sample and descriptive statistics
According to French National Institute for Statistics and Economic Studies (INSEE), the hospitality/ accommodation industry was worth 34,024 companies in 2009\(^ {11} \), accounting for approximately 1.18% of the total amount of companies in the whole French industry. In addition,

\(^{10}\) Non-bankrupt firms
\(^{11}\) Source INSEE – Esane (2009), Principales caractéristiques des entreprises en 2009
turnovers represented 20,312 million Euros and counted for 0.59% of the French turnovers in the industry. The hospitality industry with approximately 228,263 employees counted for 1.5% of the employees in the French industry\textsuperscript{12}.

All the limitations previously mentioned are applied in order to obtain the best homogeneous data. From these limitations, are exhaustively gathered, calculated and obtained 296,450 initial observations. Among the whole accommodation industry, this study concerns approximately 15,567 number of companies\textsuperscript{13}, which is about half of this specific industry. Moreover, remaining firms are particularly rather small and account for almost the other half\textsuperscript{14}. On the side of big corporations and multinationals, only 144 firms are not included in this study. Therefore, this study targets the core population of the hospitality/accommodation industry.

After defining all the previously presented data specificities, a big database and sample of firms is obtained. The major difficulty is to gather bankrupt firms that are rare compared to non-bankrupt firms. Therefore, the study and the firms to be studied are dependent on the number of bankrupt firms. A sample of 227 bankrupt companies’ observations is finally obtained and matched with an equivalent sample observations’ size of 227 non-bankrupt firms.

A stratified random basis without replacement is performed to obtain equal group with the same proportions, because non-bankrupt firms are more abundant than bankrupt firms in the initial sample. In addition, the fact to set data from opposing groups in equal proportion strengthens the robustness of the developed model against the loss of the assumptions of multivariate normality and homoscedasticity (Bardos, 2001).

The model to develop will attempt to detect bankruptcies three years in advance. Therefore, a year sample is composed of four different profiles of firms: (i) companies that went bankrupt one year later from the year studied, (ii) companies that went bankrupt two years later from the reference year, (iii) companies that went bankrupt three years later from the reference year, and in the last case (iv) companies that did not go bankrupt within the three years’ period studied. For example, considering the 2003 year samples (see Table 1), four different profiles of companies are included. From year 2003, three firms went bankrupt a year later – in 2004; 12 firms went

\textsuperscript{12} Source INSEE – Suse (2008), Principaux indicateurs économiques et comptables par secteur
\textsuperscript{13} Source Bureau van Dijk – Orbis (2009)
\textsuperscript{14} Turnovers inferior than 100k€ and specific under the one defined in the precedent part
bankrupt two years later – in 2005; 11 firms went bankrupt three years later – in 2006; and that 26 firms did not go bankrupt within at least the three years long horizon studied (sample 2003). Table 1 depicts the different years’ samples and their composition that are used in this study.

Table 1: Number of bankrupt and non-bankrupt companies classified on their reporting year and per horizon of bankruptcy (if applicable)

<table>
<thead>
<tr>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>12</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>29</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>22</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>23</td>
<td>26</td>
<td>18</td>
<td>7</td>
<td>31</td>
<td>59</td>
<td>40</td>
<td>11</td>
<td>all NB</td>
</tr>
</tbody>
</table>

B: Bankrupt; NB: Non-bankrupt

The database obtained as depicted in Table 1 and Table 2 is divided into two almost equal sample proportions (50/50): one as training data and the other as testing data. The training data is set on periods from 2001 to 2006 inclusively. The testing data is set from 2007 to 2009 inclusively.

Table 2: Number of bankruptcy and non-bankruptcy companies on a 3 years period horizon

<table>
<thead>
<tr>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>18</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>12</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>29</td>
<td>22</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>11</td>
<td>2</td>
<td>2</td>
<td>28</td>
<td>27</td>
<td>-</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>23</td>
<td>26</td>
<td>18</td>
<td>7</td>
<td>31</td>
<td>59</td>
<td>40</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>24</td>
<td>46</td>
<td>52</td>
<td>36</td>
<td>14</td>
<td>62</td>
<td>118</td>
<td>80</td>
<td>22</td>
<td>Total</td>
</tr>
</tbody>
</table>

Source: Bureau van Dijk – Orbis (2011)

In Table 2, K= 1 – depicts the number of firms that bankrupt one year from the reference studied; K= 2 – depicts the number of firms that bankrupt two years from the reference studied; K= 3 – depicts the number of firms that bankrupt three years from the reference studied; and K= 4 depicts the number of firms that do not bankrupt within at least the three years long period studied of the sample reference.
As it can be seen in Table 2, samples for three years are not fully completed. Some data are missing in years: 2001, 2008, and 2009. Missing values are marked with: “–”. For 2001, there are missing values for bankrupt firms at a year one horizon (K= 1). In this last case, data are not available in the database. For 2008, data are missing for bankrupt firms at a horizon of three years ahead of the actual reporting year – 2008. This can be explained by the fact that data as financial statement information often do take time to be collected (~ 1 year), and therefore they are not available at the time of collection in this study for the last period (K= 3). For the same reason in 2009, data numbering bankrupt firms are not available for two and three years from 2009 (K= 2 and K= 3).

After collection, data need to be reprocessed to keep an economic sense and to match the method’s characteristics. The next part presents this work.

5.2. Data reprocessing

Bankruptcy studies of firms show fluctuating results from one extreme to another. Therefore, it is necessary to reprocess some data in order to make them feet with the method’s assumptions and constraints. One of them is the monotony of the ratios used. The monotony means that: the highest the ratio’s value is – the best its financial situation is; or depending on the situation: the highest the value of the ratio is – the worst the financial situation is.

However, it can be difficult to find ratios exclusively satisfying these properties. Thus, for certain ratios a reprocessing is sometimes necessary. For example, some ratios can have their denominator negative. Others can have both their numerator and denominator negative. This type of ratios may be difficult to analyze, but important ratios to describe the process of a bankruptcy may sometimes lie in these difficult categories. Therefore, it becomes important to reprocess information and data. To this end, several techniques are applied to adjust data and the most important are presented below.

5.2.1. Missing observations and imputation

Missing observations are data that are not available in the dataset. The non-availability of data can be total or partial. Total is when all observations for a firm are missing and partial is when few observations for a firm are missing. Missing observations can be corrected according to
different techniques one of which is imputation (substituting missing values by the average or a certain quantile). However, in the special case of bankrupt firms that often have atypical data due to exceptional financial condition (bankruptcy), it is very difficult to reprocess this type of data. In practice, in previous studies the reprocessing of missing observations did not particularly improve the results from the discrimination (Bardos, 2001). Therefore, it is preferable not to reprocess or substitute missing observations and to let the dataset free of this type of adjustment.

5.2.2. Extreme values and winsorization

Extreme values are data that are shifted from the center or the mean of group observations, such as it is often the case with bankrupt data. These types of data influence averages and covariance matrices that are directly involved in the development of the model (LDA). Therefore it is necessary to reprocess these values, for which different techniques are applicable. One of these techniques – the winsorization of extreme values, is applied in this study.

The principle of winsorizing ratios is to limit extreme values in a data distribution by substituting and transforming its extreme values. For example, within a distribution of observations, extreme values of distribution queues can be transformed within the determined statistical boundaries (i.e. 1%, 5% or 10% quantiles). Winsorization is carefully applied to the variables in this study to preserve the statistical and economic information of the data. In general, the winsorization is performed at the 99% quantiles for the upper statistic boundary, and at the 1% quantiles for the lower statistic boundary.

5.2.3. Misleading values

In the previous part, classic adjustments of ratios are presented to describe either missing or extreme data. Therefore, data are left unprocessed or are substituted by boundaries’ values.

For misleading values, the situation is a bit different and somehow more problematic because without careful testing of the dataset, it would be almost impossible to detect this profile of incorrect values. In this study, two types of misleading values are identified by testing the input signs: (i) ratios that have the denominator negative or null, and (ii) ratios that have both numerator and denominator negative or null.
**Ratios that have the denominator negative or null**

The first case of misleading values encountered, is for ratios for which the denominator could become negative or null, therefore making the data no longer monotonous. Indeed, ratios having their denominator negative or null are losing their monotony, which is a pre-requisite for the linear discriminant analysis and therefore, needs to be adjusted.

For example, empirically the cash-flow to short-term debt ratio appears to have sometime its denominator equal to zero. Therefore, in order to keep a maximum of observations to develop the model, it is necessary to reprocess misleading observations.

Depending on the value of the numerator when the denominator is equal to zero, two possible reprocessing are conducted. First situation, cash-flow is positive and short-term debt is equal zero. In a situation where having cash-flow and no debt is a potentially good sign, the value of the superior boundary is attributed to the data. In the second situation, where both the numerator and the denominator are equal to zero (more rare), data are attributed the value of the inferior boundary, because, having no cash-flow and no debt is potentially a bad sign.

In addition, two other situations appear when short-term debt has a positive value. First, when the cash-flows are equal to zero, the value of the inferior boundary is attributed to the ratio as a reprocessing (see situation in top corner left of Table 3). In this last situation, having no cash-flow and short-term debt is potentially not a good sign, which justifies the inferior boundary as a reprocessing. Finally, for the remaining and the most common situation – when both the numerator and the denominator are positive depending on the value of the ratio, it is appropriate to apply potential winsorization within the lower and the upper boundaries of the distribution.

Table 3 below depicts all the discussed situations.

**Table 3: Example of adjustment when the denominator is negative or null (Cash-flow to short-term debt example)**

<table>
<thead>
<tr>
<th>Cash-flow</th>
<th>Short-term debt</th>
<th>0</th>
<th>+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inferior boundary</td>
<td>Inferior boundary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superior boundary</td>
<td>Potential winsorization within limits [inf., sup]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Ratios that have both numerator and denominator negative or null**

A second case of misleading values is identified. This last situation is even more difficult than the first one because ratios can simultaneously have their numerator and denominator negative. Reprocessing need to be conducted because data are no longer monotonous and lose their economic sense. However, empirically this latest situation is not abundantly encountered (at least on the 26 short listed discriminant variables – see Appendix B).

Appendix D presents a description of the reprocessing that are performed in this study on 26 short listed discriminant ratios (when applicable). Type 1a, 1b, and 1c examples depict situations related to the first profile of misleading values (only denominator negative or null). Type 2a and 2b examples depict situations related to the second profile of misleading values (both numerator and denominator could become negative).

Finally, among the variables that are potentially misleading values and need special reprocessing, only cash flow to short-term debt ratio is retained within the final model. It means that the most of the variables included in the final model, except cash flow to short-term debt ratio, are just slightly reprocessed within boundaries (at 1% and 99% quantiles).

Data once prepared and reprocessed are used in the model development. The next section presents this step.
6. Model development

Once data are prepared, the next important part is the development of the final model and the selection of discriminatory variables that go in it. From the 74 ratios initially selected for this research (Appendix A), a short list of 26 discriminant ratios is selected (Appendix B). Then, from several combinations of the 26 discriminant ratios, few final models are proposed for validation.

6.1. Selection of discriminant variables

Bankruptcy prediction is a challenging exercise. Due to the multitude of factors that influence the bankruptcy’s process, discriminations or differences between groups are difficult to determine. In particular, because of the nature of the data (financial statement data), discrimination will not be clear in the most of the cases\textsuperscript{15}. Therefore, small oppositions and differences between groups are satisfying. More importantly, because of the assumption of multinormality, distributions that follow a Gaussian distribution are preferred, or at least approached due to the frequent violation of this latest assumption in LDA. However, the model is very robust to the loss of the assumption of multinormality as long as distributions between opposing groups approach a Gaussian distribution (Bardos, 2001).

The selection of discriminant variables is organized into the following three part: at first ratios’ distributions are compared through histograms and testing, then oppositions between groups are examined through quantiles distribution tables and testing, and finally some tests of variables correlations are performed.

6.1.1. Analysis of histograms distribution and testing

In attempt to discover ratios that are discriminant between the two opposite groups, histograms of the descriptive variables are constructed. Histograms offer two major advantages in attempt to develop the best model. Their analysis is more robust to extreme values than the simple analysis

\textsuperscript{15} There is evidence in the literature that financial ratio variables generally exhibit non-normal distributions (Balcaen & Ooghe, 2006)
of averages and standard deviations. In addition, histograms permit the graphical verification of
the monotonous property (i.e. if a ratio increases, the risk decreases or vice-versa).

It is difficult to create a clear opposition between groups, because variables are complex to
interpret. Four different cases or situations can be depicted to describe comparison between
groups. The ideal situation is shown in Figure 2, where there is a perfectly clear linear
discrimination between the two groups: non-bankrupt and bankrupt firms. This is the only
situation in which variables are selected for further consideration in the model development.

**Figure 2: Clear discrimination between opposing groups**

However, in practice, clear discrimination as it appears in Figure 2, is rare. Nevertheless, selected
variables will still attempt to conform with the assumptions of the model and with the graph
depicted in Figure 2. Variables with opposing groups that are not conform with the assumptions
of the model and with the form presented in Figure 2, are not considered further in this study.

Figure 3 and Figure 4 show examples of discriminations that are not linear. Consequently,
variables in this situation are not selected for a linear discriminant analysis, but however could be
selected in a nonlinear discriminatory method such as neural networks or decision trees
methods.
Finally, Figure 5 presents the last situation when there is no discrimination at all. Ratios in this situation are not considered further in this study.

In addition to the histograms analysis, Kolmogorov-Smirnov tests are performed to decide whether or not the distribution of a variable is the same across different groups as the null
hypothesis. The results from histograms and Kolmogorov-Smirnov tests for the final discriminatory model are presented in the Appendix F. Results demonstrate differences between opposing groups: bankrupt and non-bankrupt firms, on the variables selected in the final model.

6.1.2. Analysis of quantiles distribution and testing

The analysis of the quantiles tables are a good addition to the histogram analysis. Compared to the previous graphical investigation, quantiles tables are more precise to analyze. The principle is to look at a comparison of quantiles between the two opposing groups. If significant differences are found, variables may be further considered in this study. In particular, it is very interesting to compare opposing groups’ results between their first and third quartiles (Q1 and Q3), where half of the observations are numbered. This process can also help in detecting some potential asymmetry. For example, an opposing group may be only discriminant for extreme or high values, or vice versa.

In addition to the quantiles’ analysis, non-parametric tests of quartiles are performed. The median test is performed, which tests if the medians of the population of the two samples groups are identical. Specifically, data from the compared samples are assigned to two groups. One group is compared to higher values than the median value of the two groups combined. The other group is composed of values that are less or equal to the median. Median scores equal 1 for observations greater than the median and 0 otherwise, such as:

\[
a(R_j) = \begin{cases} 
1 & \text{if } R_j > \frac{n + 1}{2} \\
0 & \text{if } R_j \leq \frac{n + 1}{2}
\end{cases}
\]

Results from the quantiles tables and the median tests for the final model developed are presented in the Appendix F. They all confirm differences between bankrupt and non-bankrupt groups on the variables selected.

16 In addition, Appendix E presents an example of the SAS coding used to adjust the histograms’ frequencies
6.1.3. Correlation analysis

Correlation tests of variables are performed in each economic theme (i.e. liquidity, profitability, etc.), in order to distinguish variables that are too much correlated in the same economic theme, and therefore, should not be considered together in the final multivariate model.

Two techniques are used to perform this analysis: Pearson and Spearman analysis. The Pearson analysis is sensible to extreme values and the Spearman correlation is better to detect atypical values (Stili, 2002). Results from Spearman and Pearson’s correlations for the final variables selected in the model are presented in the Appendix F. In general, correlations’ results are relatively low between variables, meaning that data cover different parts of the business without overlapping.

The previously described selection is solely univariate and do not consider multivariate associations of the ratios. In the following part, several models are constructed on the basis of their best association to discriminate bankrupt to non-bankrupt firms.

6.2. Combination of variables to form the discriminatory model

In the previous part, discriminant variables were selected and conducted to 26 final selected ratios (see Appendix B for a list). The next challenge is to associate these 26 ratios in a multivariate model in order to obtain the ultimate discriminatory model. Since this process is mainly iterative, there is no claim regarding the optimality of the determined discriminant model (E. I. Altman & Edith Hotchkiss, 2006). However, the function developed aims to be the best among the ones tested. In order to develop different sets of discriminant models, different approaches are utilized. All models are developed on the training data.

A stepwise selection is performed first (PROC STEPDISC SAS). The principle is to select at a different step, variables that contribute the most to the discriminatory power of the model as measured by Wilks’ lambda. However, there are two drawbacks in using this method. At first, the stepwise selection assumes that variables are multivariate normal and with common covariance matrix, such as for the LDA. Nevertheless, using this technique should be robust to a certain loss of these assumptions (Bardos, 2001). The other drawback is that in the selection of variables for entry, the stepwise process selects only one variable at a time to be entered in the model and does
not consider relationships between variables that have not yet been selected. Therefore, some important variables could be excluded from this process.

Considering the advantages and the drawbacks from the stepwise selection, a combination of different approaches is a valuable help in selecting variables for the discriminant model. Thus, decision trees are performed in order to distinguish strong discriminant variables and to detect certain behaviors. This process offers an interesting exploratory analysis of variables. Then, correlation of variables, economic sense and knowledge of the data, subjective judgment and a combination of all approaches are used to help find the best combination of discriminant variables.

All these processes are performed to meet at best the following six characteristics: (i) the model considers most of the different economic aspects of a firm (liquidity, profitability, etc.), (ii) the model discriminates jointly and efficiently groups to oppose, (iii) variables should be independent between each other and in particular to be linearly independent between each other, (iv) variables should be monotonous, (v) multivariate normal, and (vi) homoscedastic. The two latest being robust to a certain degree of the loss of assumption, as mentioned previously.

In each economic theme (liquidity, profitability, etc.), different variables appear to be discriminant, which makes hundreds of possible combinations of the final model (see Appendix B for a list of discriminatory variables by economic theme). However, variables from the same economic theme should not be associated in the same model in order to avoid information redundancy and in particular to avoid multicollinearity that leads to misleading and opposite meaning.

Finally, different alternative models are developed according to the previously enumerated approaches and goal characteristics. Few models of interest are established, out of which one final model is selected for the research study. The next section presents the results of the final model selected.
7. **Empirical results**

From the previously mentioned work, different models of interests are developed through the PROC DISCRIM in SAS. After testing how the models performed on the training data, few last models are kept for final consideration on the testing data. Thus, only one last model is kept as the discriminant model because of its overall better performances and characteristics.

It is important to present detailed results because they permit to understand the limits and strengths of the developed model, for a better utilization. The performances achieved by the final model during each steps of the selection process are presented in the following four parts: (i) to begin, the results from the validation of a function estimate are presented. The idea in this latest part is to select within all models the best estimate based on the training sample. (ii) Then, the results from the validation of the final function are presented. At this stage, the best model is selected. Other good performing models are kept in reserve in case a last minute problem of the main model may appear in the remaining two stages. (iii) Then, the cut-off value is optimized according to subjectively chosen criteria, by nonlinear programming that enhances the model’s performance. (iv) Finally, the optimized model is tested against the training data. Thus, if passed, the model is checked on the testing data.

7.1. **Validation of a function estimate**

The validation of a function estimate is the first step performed in the validation and confirmation of the final model. At this stage, different year estimates of models developed are compared for each model to determine what year estimate is the best as well as to check for the overall robustness to economic changes of a model in general. Precisely, a good selected function should have the two qualities: (i) a good rate of good classification and (ii) the stability of its performance over time. Therefore, the selected function should permit discrimination for several years after implementation.

Different estimates are built and tested on the training data. In particular, for each model five estimates are obtained and tested, almost one for each year sample of the training data (except 2001). Each estimate (i.e. 2002, 2003, etc.) is composed of firms that bankrupt in one year’s
horizon, two years’ horizon, three years’ horizon, and of firms that never bankrupt in the three years long period. This situation was previously depicted in Table 1 and 2 in the data section.

In the case of the final model, the “2003 estimate” is selected for further consideration in the validation of the estimated function’s part. Its results and comments are presented below.

From the figures depicted in Appendix G, which represent different estimates’ performances of the final model, it can be concluded that the final model and its different estimates are performing correctly within the period of study. However, two estimates (2004 and 2005) do not demonstrate results as good as the others. Regarding the 2005 estimate, the results are not high, especially for the rate of good classification of bankrupt firms; this result is explained by the small amount of data \(^{17}\) that are used to develop this model based on 2005 estimate.

For the 2004 estimate: results are not as high as for the other estimates. In this situation, 36 data are utilized to develop this model based 2004 estimate, which even if not very high, should be sufficient to develop a correct model. Therefore, the cause should not be the amount of data. In addition, the most striking fact is the wider spread that exists between bankrupt and non-bankrupt firms on this estimate. This can be interpreted as a good sign, since it may be corrected by a change in the cut-off value that would readjust and balance rate of good classification between the two groups studied. However, this cut-off change may be good for the 2004 estimate results but not for the other estimates. Therefore, in both cases – 2004 and 2005, estimates are eliminated from the study.

Still, in the case of the final model, three last estimates remain at this stage of the process: 2002, 2003, and 2006 estimates. The results of these estimates show that they are close to each other as shown in Appendix G. However, these estimates also have their own characteristics. The 2006 estimate displays a lower rate of good classification on average compared to the two others (2002 and 2003). However, the 2006 estimate produces the closest results for both groups – bankrupt and non-bankrupt. The model based on the 2002 estimate demonstrates on average a better percentage of good classification than the 2006 estimate. However, results from opposed groups on this estimate demonstrate some disparities – in particular in year 2003 and 2004 (see top corner left of the figure “estimate 2002” in Appendix G).

---

\(^{17}\) Only 14 firms in the 2005 estimate sample, see Table 1 and 2 in the data section
Preference is given to the 2003 estimate. It is selected because of its ability to combine better aspects from the 2002 and 2006 estimates. Also, for these three last estimates (2002, 2003, and 2006), the year 2005 is not over considered in the comparison, because, as previously mentioned, there are solely 14 firms composing this particular year. Therefore, results for year 2005 should be considered carefully.

Among other models with their best estimates, this 2003 estimate is considered for further testing in the following part.

### 7.2. Validation of the final function

Among various models, only the best estimates are selected. This selection process helps in assessing the overall behavior of the models. Therefore, the model that is considered in this part: the validation of the final function, demonstrates a very good performance.

In order to select one model, a final validation of the function is performed on all models. The idea is to investigate in detail the results from the few remaining models of the whole selection process. Among other figures, confusion matrix and graphical plots are used to help determine the best model. The results and the reasons to select the final model are demonstrated in the following paragraphs.

Starting with classification charts or confusion matrices is a good element to help in distinguishing which final model is the best to select. Table 4 presents the results of the selected function.

<table>
<thead>
<tr>
<th>True classes</th>
<th>Predictions</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bankrupt</td>
<td>Non-bankrupt</td>
<td>Sum</td>
<td></td>
</tr>
<tr>
<td>Bankrupt</td>
<td>77.78%</td>
<td>22.22%</td>
<td>100.00%</td>
<td></td>
</tr>
<tr>
<td>Non-bankrupt</td>
<td>10.26%</td>
<td>89.74%</td>
<td>100.00%</td>
<td></td>
</tr>
<tr>
<td><strong>Percentage correctly predicted</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>83.76%</strong></td>
</tr>
</tbody>
</table>

Table 4: Confusion matrix of the training data for the selected model based on the 2003 estimate
The results of the selected function on the training data are promising with an overall percentage of correct prediction equal to 83.76%. However, two remarks should be taken into consideration: first, even if results are promising\textsuperscript{18} they should be confirmed on a testing data (performed later in this study), the second remark concerns the particular rates of good classification of non-bankrupt and bankrupt firms. Indeed, when attention is paid to these rates, it can be concluded that they are unbalanced and that there is an important spread of 11.96% between their performances. Seen from another angle, type I (22.22%) and type II (10.26%) errors are too different from each other.

Type I errors represent the misclassification of bankrupt firms as non-bankrupt ones. Type II errors are the misclassification of non-bankrupt firms as bankrupt firms. Therefore, the model appears to be biased in favor of non-bankrupt prediction and to the detriment of bankrupt firms, which are the ones of particular interest. However, as it is demonstrated later in this study, this apparent problem can be fixed by modifying the cut-off value.

In addition to confusion matrices, it is interesting to investigate further the results of the different models on different supports as shown in Figure 6 and Table 5. The graphical investigation in Figure 6\textsuperscript{19}, depicts the rates of good classification for the whole period studied. The final model shows a high percentage of correct classification with results for both groups that look close.

However, the confusion matrix (Table 4), shows that an important spread between results from both groups exists. Table 5 helps investigate further this fact. It can be seen that globally the rate of good classification remains above 70% for both groups. It confirms as well that there is a spread in favor of non-bankrupt firms, which demonstrates a problem of type I error.

\textsuperscript{18} It is usually difficult to obtain high percentages of good classification, see Kim (2011) for a list of results that are usually found

\textsuperscript{19} Note that Figure 6 is the same as the 2003 estimate depicted in Appendix G
Except the type I error, this model has an overall rate of good classification that looks relatively stable for the entire period of the study. It becomes interesting to investigate further the potential of this model. This can be done by looking at the different rates of good classification presented by a horizon of bankruptcy. For example, this study showed earlier that a year sample (i.e. 2003) is composed of firms that would bankrupt one, two, and three years from the stated reporting year\(^{20}\). These different horizons of bankruptcy are presented in the rest of this study as H1 (stand for one year prior bankruptcy), H2 (stand for two years prior bankruptcy), and H3 (stand for three years prior bankruptcy). Figure 7 presents rates of good classification per horizon of bankruptcy.

\(^{20}\) Also were included firms that would not go bankruptcy, but are not considered worth mentioning and studying in this special case
From Figure 7, it is obvious that the relative stability of the rates of good classification previously mentioned is under question. However, this instability may be due to few firms that are attributed to each category. As it can be seen from Table 6, when data are split in different horizons of bankruptcy, only few firms are attributed to each category, which may alter or lessen the results of each rate of good classification.

Finally, one last but not least remark can be drawn from Figure 7 and Table 6 regarding the average percentages of rates of good classification on the training period. Both figures show that H2 firms are classified at 83.87%, H1 are classified at 82.35%, and H3 are classified at 73.91%. Therefore, on average H2 firms are better classified than H1 and H3 firms.
In the case of H3 firms this conclusion is logic due to the fact that their bankruptcy in about three years is more difficult to predict. However, it is surprising to see H2 firms better classified than H1 firms. Hence, the two following remarks:

The first remark, concerns the nature of the model. As presented above, firms that will go bankruptcy in two years are easier to predict than firms that will go bankruptcy in one year. It would be better to have a model that can forecast better the bankruptcy of H1 than the one of H2. However, this is a difficult task and can be explained by the two following reasons: (i) first, when approaching bankruptcy firms and managers often do not provide their financial results to related parties. Therefore, the H1 period often has less firms than the two other categories. (ii) The second reason is that firms when they approach bankruptcy are very volatile and provide very extreme results. These two last facts contribute to the difficulty of forecasting bankruptcies.

The second remark concerns the previously mentioned problem of type I error. Seen from this angle, type I error in the case of H1 and H2 firms does not represent a major difference any longer. Firms that will go bankruptcy within two years (H1 and H2) have relatively closer rate of good classification than non-bankrupt firms (Table 4).

Taking into account all the conclusions drawn so far for this model it may be sufficient to stop here and to go with this model. However, a further investigation of the potential of the studied model showed some problems or particularities, which represent an opportunity to better understand the model.

It is essential to improve this model by including some more balanced traits and features such as a more balanced rate of good classification between both groups (or closer type I and II errors), and to increase the predictability of firms that will go bankruptcy in about one year horizon (H1). As this possibility exists, it appears interesting to optimize and improve this model with some nonlinear programming as presented in the following part.

7.3. Optimizing the cut-off value through nonlinear programming

It is generally agreed upon that type I errors are more costly than type II errors for several reasons including: loss of business (i.e. accounting auditor to decide whether or not an ongoing concern should be applied); damage to a firm’s reputation; and potential lawsuits and court costs (Altman,
Haldeman, & Narayanan, 1977). Due to the subjectivity of these misclassification costs, in practice, most researchers minimize the total error rate and therefore, implicitly assume equal misclassification costs (Balcaen & Ooghe, 2006). Some researchers and practitioners have optimized the type I or type II classification errors through the Gini coefficient (Ooghe & Balcaen, 2003), the R-square measures and likelihood or measures based on entropy (Zavgren, 1985).

As the precedent part demonstrates, there is an unbalance between type I and type II errors. This can be solved by a modification of the cut-off score. Indeed, depending on what value is attributed to the cut-off, a discrimination plane selects more firms from group 1 or 2 (bankrupt or non-bankrupt firms).

Initially, the cut-off score is set to 0. The basic principle is that the higher the cut-off score, the more it discriminates the bankrupt firms. Per opposition, the lower the cut-off score, the least it discriminates the bankrupt firms. Figure 8 below depicts the initial situation of the distribution scores when the cut-off value is still equal to 0. In addition to the previous remarks on the model strengths and weaknesses, Figure 8 demonstrates that it may be possible to increase the results’ model and discrimination by shifting the cut-off value to the right.

**Figure 8: Frequencies and scores of the opposing two groups before cut-off optimization**
Therefore, the modification of the cut-off score constitutes an additional advantage of the linear discriminant model for increased results. It is possible to optimize the cut-off value through nonlinear programming in accordance with subjectively chosen objectives and constraints. In this study, it seems interesting to maximize the overall rate of good classification on the training data, which equals nonlinear programming language by minimizing the overall rate of misclassification error. And this is subject to constraints that both groups (non-bankrupt and bankrupt firms) must have a good classification rate on the training data set (for example: above 85%).

Hence, the nonlinear programming problem can be stated such as in Equation (7.1).

\[
\begin{align*}
\text{Max } R_{tot} &= 0.5r_{NB} + 0.5r_B \\
\text{Subject to:} & \\
r_{NB} &> 0.85 \\
r_B &> 0.85
\end{align*}
\]  

(7.1)

where \( R_{tot} \) is the overall rate of good classification to maximize, \( r_{NB} \) is the percentage of good classification of non-bankrupt firms, and \( r_B \) is the percentage of good classification of bankrupt firms.

Because nonlinear problems are very complex, it is not always easy to find optimal solutions, which is the case of the findings here. However, this whole process helps a lot in finding a very good and feasible solution. Therefore, according to the problem stated in Equation (7.1), a feasible solution for the cut-off score equals to: 0.24809631. It will be referred to as 0.25 for easy reference.

With a cut-off value of approximately 0.25, the previous hypothesis on the distribution of opposing groups that shift to the right is verified. Indeed, it is possible to improve the discrimination between both groups by changing the cut-off value. Figure 9 depicts this new situation as opposed to Figure 8.
As a final remark, the modification of the cut-off score is of great interest for practitioners as it provides them with more flexibility to tailor this tool to their needs. Depending on the goal targeted, it is possible to set as another objective the highest possible rate of good discrimination of bankrupt firms. For example, if someone looks for the healthiest firms or is very selective or risk averse, a potential solution would be to set the cut-off value as high in order to select solely non-bankrupt firms. Looking back at Figure 8\textsuperscript{21}, a feasible solution to the latest case would be to set a cut-off value at approximately 4. Using the same reasoning in the opposite sense, it would be possible to exclude firms that have few chances of survival at a certain degree of risk or percentage of misclassification (i.e. to set a cut-off value at -10 or -5 and so forth (see Figure 8)).

The results from Equation (7.1) and the optimization problem are presented in the next part.

### 7.4. Optimized results

After final investigation and optimization of the model, results from the scoring function on the training data appear higher than before optimization. The confusion matrix depicted in Table 7 shows that on the training data the overall rate of good classification has increased a little from 83.76\% (Table 4) to 84.62\% (Table 7). More importantly, the two opposing groups are now

\textsuperscript{21} Situation before optimization
classified in similar proportions. Indeed, type I (16.24%) and type II (14.53%) errors are now only separated by a spread of 1.71% compared to the previous 11.96% (Table 4). This new spread on the training data is almost seven times lower than before. Only the specific rate of good classification for non-bankrupt has decreased, which is a direct consequence of adjusting the cut-off value. However, this rate is still largely acceptable and is highly performing on the training data.

Table 7: Confusion matrix for the cut-off optimized model per years

<table>
<thead>
<tr>
<th>True classes</th>
<th>Predictions</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bankrupt</td>
<td>Non-bankrupt</td>
<td>Sum</td>
</tr>
<tr>
<td>Bankrupt</td>
<td>83.76%</td>
<td>16.24%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Non-bankrupt</td>
<td>14.53%</td>
<td>85.47%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

| Percentage correctly predicted | 84.62% |

The good test results of the model on the training data need as a final proof to be tested on the testing data that was not previously used in the selection of any model. The idea of having a testing data is to test one last time the model with a new data in order to check if the model is not biased by being used on the training data. For example, the model could have a high rate of good classification on the training data because it would be too specific to the features and traits of this sample. However, the model when put in a new situation such as a new economic environment would no longer function and discriminate as it should.

Therefore, the model is tested on a testing data that accounts for almost 50% of the initial data and is based on years 2007 to 2009 inclusively. The results from these tests are shown in the confusion matrix that follows.
Table 8: Confusion matrix for the optimized cut-off model on the testing data

<table>
<thead>
<tr>
<th>True class</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bankrupt</td>
</tr>
<tr>
<td>Bankrupt</td>
<td>87.27%</td>
</tr>
<tr>
<td>Non-bankrupt</td>
<td>14.55%</td>
</tr>
</tbody>
</table>

Percentage correctly predicted 86.36%

Table 8 shows that the results of the testing data are very promising. In this difficult period where there is an increase of bankrupt firms 22, the model performs even better than before with 86.36% of overall rate of good classification. This increase in overall rate of good classification is particularly due to an increase in bankruptcy predictions (87.27%) or to a decrease of type I errors (12.73%). Non-bankruptcy predictions on their side remain the same in the training and in the testing data. Therefore, from the three confusion matrices depicted before (Table 4, Table 7, and Table 8), overall results of the model are demonstrated as performing and progressing well.

A further investigation of the results from the opposing two groups is performed. Figure 10 shows the rates of good classification for the whole data. It is obvious that both groups – non-bankrupt and bankrupt firms, have similar rates of good classification: almost always above 80%. Like previously, results from year 2005 should not be considered seriously because of the rather small amount of data that are available for this year (see No. firms in Table 9).

---

22 Period corresponding to the financial crisis of 2007
In addition, Table 9 shows that the rates of good classification for both groups are relatively stable over time.

<table>
<thead>
<tr>
<th></th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-bankrupt</strong></td>
<td>(%)</td>
<td>(No.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>firms</td>
<td>83.33</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>82.61</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>92.31</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>83.33</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>57.14</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>90.32</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>83.05</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>87.50</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>90.91</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bankrupt</strong></td>
<td>(%)</td>
<td>(No.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>firms</td>
<td>75.00</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>86.96</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>84.62</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>77.78</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>100.00</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>83.87</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>86.44</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>87.50</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>90.91</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As a final investigation and again as a matter of understanding better the potential of the model it is interesting to perform the detailed results of the model per different horizon of default.

It can be seen form Figure 11 that different horizons of bankruptcy are less volatile than before optimization (vs. Figure 7). However, two peaks are still noticeable in 2004 and 2007. First – in 2004, the noticeable peak is for H3 that has a classification rate of approximately 50%. This low classification rate should be contextualized by the fact that only two firms are considered in this portion (see No. of firms in Table 10). For the other peak – in 2006, the situation is attributed to
H1 firms. In this case the model could detect some pre-crisis effects. However, again results from this portion are diminished by the fact that only three firms’ data are available in this portion.

Figure 11: Rate of good classification for the cut-off optimized model for bankrupt firms decomposed per horizon of default for the whole period

Table 10: Rate of good classification for the cut-off optimized model explained per horizon of bankruptcy

<table>
<thead>
<tr>
<th>Year</th>
<th>H1 (%)</th>
<th>H2 (%)</th>
<th>H3 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>100.00</td>
<td>66.67</td>
<td>77.78</td>
</tr>
<tr>
<td>2002</td>
<td>66.67</td>
<td>80.00</td>
<td>83.33</td>
</tr>
<tr>
<td>2003</td>
<td>77.78</td>
<td>82.35</td>
<td>81.82</td>
</tr>
<tr>
<td>2004</td>
<td>100.00</td>
<td>85.71</td>
<td>50.00</td>
</tr>
<tr>
<td>2005</td>
<td>88.89</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>2006</td>
<td>90.91</td>
<td>89.66</td>
<td>82.14</td>
</tr>
<tr>
<td>2007</td>
<td>90.91</td>
<td>81.82</td>
<td>85.19</td>
</tr>
<tr>
<td>2008</td>
<td>90.91</td>
<td>82.14</td>
<td>85.19</td>
</tr>
<tr>
<td>2009</td>
<td>90.91</td>
<td>82.14</td>
<td>85.19</td>
</tr>
</tbody>
</table>

Finally, after all these promising findings, it appears necessary to check the robustness of the developed model. The next section presents a robustness analysis of the model.

---

23 Refers to the financial crisis of 2007-2008
8. Robustness analysis

Now that the final model is developed, optimized and validated on both training and testing data, it is interesting to investigate further the properties of the model. Different analyses and tests are performed in this section. To begin, tests on the major model’s assumptions are performed. Then, broader tests of the model good performance are implemented.

8.1. Tests of the assumptions

To control the robustness of the final model, tests on major underlying assumptions of LDA are performed such as: (i) test of the assumption of multivariate normality, (ii) test of the assumption of homoscedasticity, and (iii) test of the assumption of non-multicollinearity. In addition, a complete list of the assumptions of the linear discriminant analysis model is available in the Appendix H.

8.1.1. Assumption of multivariate normality

The first test performed is the test on the assumption of multivariate normality. In order to test this assumption, several univariate tests are conducted. Indeed, Looney (1995) proposed a strategy for assessing the assumption of multivariate normality that is based on commonly used tests of the univariate normality assumption. According to Looney and many other scientists a reasonable first step in assessing the assumption of multivariate normality is to test each variable separately for univariate normality. If non-normality is indicated for one or more of the variables, thus the assumption of multivariate normality can be rejected because it is known that all univariate marginal distributions of the assumption of multivariate normality distribution are themselves univariate normally distributed (Johnson & Wichern, 2007).

In particular, it is reasonable to make this approach because bankruptcy models do often violate the assumption of multivariate normality (Barnes, 1987) (S. Balcaen & Ooghe, 2006). Thus, researchers often neglect tests on assumptions. However, because of the need to understand how well the model can perform, it is necessary to check for this assumption. In addition, univariate tests produce useful insights on the variables’ behavior and strengthen their understanding.

24 Just to quote a few, see Gnanadesikan (1997) or Johnson & Wichern (2007)
A first step in assessing multivariate normality is to perform a test on univariate normality. For this reason it is essential to perform statistical tests such as the Lilliefors test and the Shapiro-Wilk test and later Q-Q plots.

**Statistical tests**
Two statistical tests are performed to asses if variables are normally distributed: the Lilliefors test and the Shapiro-Wilk test. Both have for null hypothesis that a data are normally distributed. Table 11 depicts the result outputs from both tests.

It can be seen that most of the variables are not normally distributed. In particular, none of the variables describing bankrupt firms are normally distributed at a significance of 5%. This last fact pinpoints more than ever the extreme disparity and complexity of the bankrupt data.

For bankrupt firms the situation is in general more mitigated. Ratios of Current liabilities, Debt, and Return on Assets are found to be significant at 5% level in both tests. Ratio of Asset Turnover has the assumption that normality is rejected according to the Shapiro-Wilk test, which according to many researchers (Razali & Wah, 2011) is the preferred test among those two.

Finally, the Ratio of Cash-flow to short-term Debt has the assumption that normality is rejected in both tests.

**Quantile-Quantile plot (Q-Q plot)**
Additionally to the statistical tests, the graphical method of the Normal Q-Q plot is performed. The idea is to investigate graphically as an alternative to numerical tests if a variable is distributed normally. To this end, a scatterplot is performed with the quantiles of the data studied on the horizontal axis and the expected normal results on the vertical axis. If the two distributions being compared are similar, as in the case of a normal distribution, Q-Q plot approximately lie on the 45 degree reference line (also plotted). If the compared distributions are different from the reference line, then the assumption of normality is probably violated.

From Appendix I, which depicts Q-Q plots for the final model variables, it can be seen that in the most of the cases the Q-Q plot departure is away from the reference line, which confirms the previous findings from statistical tests. Again none of the bankrupt firms’ ratios seems to be normally distributed. For non-bankrupt firms, Ratios of Debt and of Return on Assets appear to

---

25 Coded with 1
be the closest to a normal distribution. Ratios of Asset Turnover and Cash-flow to Short-term Debt are not normally distributed.

<table>
<thead>
<tr>
<th>Table 11: Tests of univariate Normality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Bankruptcy</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Kolmogorov-Smirnov</strong></td>
</tr>
<tr>
<td>Statistic    df  Sig.</td>
</tr>
<tr>
<td>Statistic    df  Sig.</td>
</tr>
<tr>
<td>Current Liabilities</td>
</tr>
<tr>
<td>Ratios          0   .157  26  .100</td>
</tr>
<tr>
<td>1   .188  26  .018</td>
</tr>
<tr>
<td>Debt Ratio              0   .129  26  .200</td>
</tr>
<tr>
<td>1   .249  26  .000</td>
</tr>
<tr>
<td>Return on Assets            0   .134  26  .200</td>
</tr>
<tr>
<td>1   .197  26  .011</td>
</tr>
<tr>
<td>Asset Turnover              0   .134  26  .200</td>
</tr>
<tr>
<td>1   .244  26  .000</td>
</tr>
<tr>
<td>Cash flow to Short-</td>
</tr>
<tr>
<td>term Debt</td>
</tr>
<tr>
<td>0   .250  26  .000</td>
</tr>
<tr>
<td>1   .336  26  .000</td>
</tr>
<tr>
<td>a. Lilliefors Significance Correction; 0: Non-bankrupt firms; 1: Bankrupt firms</td>
</tr>
</tbody>
</table>

**Multivariate test of normality**

Previous individual testings demonstrate that most of the variables in the model are not normally distributed. Consequently, there cannot be a multivariate normal distribution. Therefore, it is not necessary to further investigate this assumption and to test for multivariate normality. However, this procedure could have been conducted through the following tests: statistic measures of multivariate skewness and kurtosis (Mardia, 1970), multivariate Shapiro-Wilk test (Royston, 1983), and the Doornik-Hansen test (Doornik & Hansen, 2008).

Finally, a violation of the assumption of multivariate normality is not surprising and is more considered as the norm rather than the exception when LDA technique is applied to bankruptcy prediction. However, models tend to be very robust against this loss of assumption (Bardos, 2001). In order to limit the effect of this loss of assumption, during the selection process of the discriminant variables, it is important to select variables that are approximately normally distributed and that could be differentiated on this basis.
8.1.2. Assumption of homoscedasticity

Another important assumption is that the variance-covariance matrices are assumed to be equal. To this end, a Box’s M test is performed for which the null hypothesis is that the group variance-covariance matrices are equal. Results of this test are depicted in Table 12 and are obtained using SPSS.

| Table 12: Box's M test of equality of variance-covariance matrices |
|----------------------|-----------------|
| Box's M   | 114.596 |
| F Approx. | 6.813  |
| df1       | 15     |
| df2       | 10065.789 |
| Sig.      | .000   |

As seen in Table 12, the null-hypotheses can be rejected, which means that covariance may be different. Therefore, the assumption of equal variance-covariance matrices is rejected in this model. Again, this result is not surprising since for bankruptcy prediction models it appears to be the norm rather than the exception (S. Balcaen & Ooghe, 2006). However, the fact that the independent variables are not perfectly normally distributed may be the reason of violation of the homoscedasticity assumption. Indeed, it should be noted that the Box’s M test is very sensitive to the loss of the assumption of multivariate normality (Anderson, 2006). It is important to note that as long as the independent variables are approaching a normal distribution, the linear discriminant analysis method is robust to the loss of this latest assumption (Bardos, 2001). It becomes important to verify if the assumption of non-multicollinearity is respected.

8.1.3. Assumption of non-multicollinearity

The last important assumption verified is the non-multicollinearity of the explanatory variables. Multicollinearity appears when the explanatory variables are too correlated between each other, thus producing misleading results. It is possible to verify this assumption with the tolerance and variance inflation factor (VIF) test, while the latest is the inverse of the tolerance. These two tests are based on the results of the $R^2$ of the auxiliary regression of the explanatory variables to each other. Table 13 depicts the results of the tolerance and VIF tests.
Table 13: Tolerance and VIF for the variables in the analysis

<table>
<thead>
<tr>
<th></th>
<th>Debt Ratio</th>
<th>AT¹ ratio</th>
<th>CF-ST² Debt</th>
<th>CL³ ratio</th>
<th>ROA⁴</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tolerance</td>
<td>0.826</td>
<td>0.822</td>
<td>0.826</td>
<td>0.809</td>
<td>0.715</td>
</tr>
<tr>
<td>VIF</td>
<td>1.211</td>
<td>1.217</td>
<td>1.211</td>
<td>1.236</td>
<td>1.399</td>
</tr>
</tbody>
</table>

¹ AT : Asset Turnover; ² CF-ST : Cash-flow to Short-term Debt ratio; ³ CL : Current Liabilities ratio; ⁴ ROA : Return on Assets ratio

A rule of thumb is that if the tolerance value is smaller than 0.10 or the Variance Inflation Factor (VIF) is higher than 10, there is probably a problem with multicollinearity.

From Table 13, it is obvious that none of the computed ratios demonstrates problems with multicollinearity. Therefore, it can be concluded that the model satisfies the assumption of non-multicollinearity.

This latest finding in respect of the assumption of non-multicollinearity is probably more important than the two previous violations of assumptions since problems of multicollinearity would severely affect the results (Bardos, 2001). Therefore, this finding enhances the discriminatory power of the model.

8.2. Complementary tests of robustness

In addition to assumption tests, it is possible to test for the general good performance of the model obtained. To this end, different measures are performed such as: tests of equality of group means, logistic regression and tests of assessing the model fit.

8.2.1. Tests of equality of group means

The basic principle of tests of equality of group means is to measure if the vectors of the averages of the two opposing groups (non-bankrupt and bankrupt firms), are equal. Tests of equality of group means can be performed by the Wilks’ Lambda and F tests. Table 14 depicts their results obtained through SPSS.
Table 14: Tests of equality of group means

<table>
<thead>
<tr>
<th></th>
<th>Wilks' Lambda</th>
<th>F</th>
<th>df1</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Liabilities Ratio</td>
<td>.906</td>
<td>46.670</td>
<td>1</td>
<td>452</td>
<td>.000</td>
</tr>
<tr>
<td>Debt Ratio</td>
<td>.731</td>
<td>166.394</td>
<td>1</td>
<td>452</td>
<td>.000</td>
</tr>
<tr>
<td>Return on Assets</td>
<td>.885</td>
<td>58.603</td>
<td>1</td>
<td>452</td>
<td>.000</td>
</tr>
<tr>
<td>Asset Turnover Ratio</td>
<td>.781</td>
<td>126.674</td>
<td>1</td>
<td>452</td>
<td>.000</td>
</tr>
<tr>
<td>Cash flow to Short-term Debt</td>
<td>.876</td>
<td>63.792</td>
<td>1</td>
<td>452</td>
<td>.000</td>
</tr>
</tbody>
</table>

From Table 14, it can be seen that all variables have their Wilks’ Lambda results inferior or equal to 0.9. In this case, it means that vectors of the averages of the two groups are different. In addition, it can be seen from the F-tests and their significances (all equal to 0), that all variables present differences in their opposing groups. Therefore, results from both tests demonstrate that variables from opposing groups are different and present signs of discrimination. Thus, these findings strengthen the model developed.

8.2.2. Logistic Regression for the significance of variables

The Logistic Regression is another parametrical technique to separate groups. It presents different pros and cons compared to LDA. One of its major advantages is to offer the possibility to test the significance of the variables included in the model. Therefore, it becomes interesting to test variables of the selected model on a logistic regression in order to check how good they performed.

By definition, the logistic function has for equation:

\[ f(z) = \frac{e^z}{1 + e^z} \]

\[ = \frac{1}{1 + e^{-z}} \]  

(8.1)

where \( f(z) \) represents the probability of having a bankrupt firm conditional on the description of \( x \), such as \( P(B|x) \); \( z \) is the contribution of all independent variables, such as \( z = \beta + \alpha_1 x_{i1} + \)
\[ \cdots + \alpha_j x_{ij}, \] with the \( x_{ij} \) is the \( j^{th} \) predictor of the \( i^{th} \) case, \( \beta \) is the intercept and \( \alpha_j \) is the \( j^{th} \) coefficient, which significance is tested.

Results from the logistic regression model are computed using SAS. The outputs of the regression analysis are depicted in Table 15 below.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>5.6913</td>
<td>0.6925</td>
<td>67.5395</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Current Liabilities</td>
<td>1</td>
<td>-2.0361</td>
<td>0.5729</td>
<td>12.6288</td>
<td>0.0004</td>
</tr>
<tr>
<td>Debt Ratio</td>
<td>1</td>
<td>-3.3770</td>
<td>0.5519</td>
<td>37.4436</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Return on Assets</td>
<td>1</td>
<td>2.8584</td>
<td>1.2088</td>
<td>5.5915</td>
<td>0.0180</td>
</tr>
<tr>
<td>Asset Turnover Ratio</td>
<td>1</td>
<td>-1.1335</td>
<td>0.2086</td>
<td>29.5379</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Cash-flow to Short-term Debt</td>
<td>1</td>
<td>0.6787</td>
<td>0.4648</td>
<td>2.1323</td>
<td>0.1442</td>
</tr>
</tbody>
</table>

From Table 15, the most interesting information to consider is the significance of the variables included in the model using the Wald statistical test. It can be seen that most of the parameters are significant at a 5% level (see p values in the last column of Table 15). Only cash flow to short term debt is not significant in this regression model. In addition, the estimated signs agree with the expected signs that are found with the LDA technique (see later p. 69, Equation (10.1) and (10.2)). Therefore, the model developed previously with the LDA technique can be considered as having useful variables to describe the phenomenon studied.

8.2.3. Assessing model fit

Eigenvalues
The Eigenvalues output table computed with SPSS presents information about the efficiency of the discriminatory function. When there are two groups in the analysis, the most meaningful output to look at is the canonical correlation, which is comparable to the Pearson’s correlation between the discriminant scores and the groups. The closest the canonical correlation is to 1, the best is the model estimated. From Table 16, it is shown that the canonical correlation is relatively high, which strengthens the discriminatory power of the model.
Table 16: Eigenvalues

<table>
<thead>
<tr>
<th>Function</th>
<th>Eigenvalue</th>
<th>% of Variance</th>
<th>Cumulative %</th>
<th>Canonical Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.108</td>
<td>100.0</td>
<td>100.0</td>
<td>.725</td>
</tr>
</tbody>
</table>

Wilks’ lambda of the model

Wilks’ lambda is a measure of how well the discriminant function separates groups. The smaller Wilks’ lambda values, the better the discriminating power of the model. With a Wilks’ lambda equal to 0.474, the model appears to have a good discriminatory ability. Table 17 depicts this situation.

Table 17: Wilks' Lambda

<table>
<thead>
<tr>
<th>Test of Function(s)</th>
<th>Wilks' Lambda</th>
<th>Chi-square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.474</td>
<td>35.428</td>
<td>5</td>
<td>.000</td>
</tr>
</tbody>
</table>

Finally, the lower the significance of the model, the better its discriminatory power is. With a significance of 0 (Table 17), the model has a better chance at separating groups.

To conclude this section, the results from the robustness analysis demonstrate that the model developed has very good characteristics. With the exception of the assumptions of multivariate normality and of homoscedasticity\(^26\), other parameters demonstrate clear signs of good performance. In particular, the fact that the variables are relatively independent (non-multicollinearity) is very important. Also, the variables demonstrate signs of good discrimination (tests of equality of group means, Eigenvalues, and Wilks’ Lambda). Finally, the complementary logistic regression shows that most of the variables are found significant with same signs as the LDA.

In the following section further analyses of the model are presented.

\(^26\) Remember, these two assumptions are often violated in bankruptcy prediction models and are not an exception. In addition, models tend to be empirically robust to these losses of assumptions (Bardos, 2001).
9. Probabilities of bankruptcy and risk classes

It is important to determine an estimate for the probability of bankruptcy because it measures the risk intensity. In addition, it allows the constitution of different risk classes that identify an equivalent or a homogenous degree of risk of bankruptcies. This estimate refines the model’s classification in order to answer the second research question raised.

9.1. Probability of bankruptcy

In the case of LDA, the process leads to the discrimination of two groups in a simplistic way (solely two classification groups: bankrupt or non-bankrupt), without considering the degree of risk. Companies having a score inferior to the threshold are classified as bankrupt, whereas firms having their score above threshold are considered non-bankrupt.

There are some companies that have a very close score but are classified on the opposite sides of the threshold. These companies are considered very differently but they have a relatively closer risk. In addition, two firms that have their score far from each other but are still on the same side of the threshold, are classified as the same but essentially have a different risk profile. Therefore, in addition to the score it is necessary to add a measure of the risk intensity of each score. This can be done by the calculation of posterior probabilities of bankruptcies ($p_i$) per interval $i$ of homogenous risk of bankruptcy.

Given that the score is one-dimensional, risk classes are determined through the research of temporal stability into probabilities of bankruptcies ($p_i$). At first, companies are regrouped according to the value of their score into small intervals of amplitude 0.25 under which is computed the posterior probability of bankruptcy ($p_i$) on each intervals. Thereafter, the analysis of the $p_i$ computed allows further regrouping into homogenous intervals with similar $p_i$. Table 18 presents the scheme of the calculation of posterior probabilities.
Table 18: Organizational scheme of the calculation of the posterior probabilities

<table>
<thead>
<tr>
<th>Interval number</th>
<th>Interval score</th>
<th>Number of non-bankrupt firms</th>
<th>Number of bankrupt firms</th>
<th>Posterior probability of bankruptcy</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>$[\inf_{i}, \sup_{i}]$</td>
<td>$n_{i}^{NB}$</td>
<td>$n_{i}^{B}$</td>
<td>$p_{i}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>$n_{NB}$</td>
<td>$n_{B}$</td>
<td></td>
</tr>
</tbody>
</table>

where:

$n_{i}^{NB}$ and $n_{i}^{B}$ are respectively the number of non-bankrupt and bankrupt firms in the interval $i$.

$n_{NB}$ and $n_{B}$ are the total number of non-bankrupt and bankrupt firms, such as:

$$n_{NB} = \sum_{i=1}^{I} n_{i}^{NB} \text{ and } n_{B} = \sum_{i=1}^{I} n_{i}^{B}$$

In addition, the frequencies in percentage per categories are computed, such as:

$$f_{i}^{NB} = 100 \times \frac{n_{i}^{NB}}{n_{NB}} \text{ and } f_{i}^{B} = 100 \times \frac{n_{i}^{B}}{n_{B}}$$

$f_{i}^{NB}$ and $f_{i}^{B}$ are respectively the estimated probability of belonging to an interval $i$ conditional on the non-bankrupt or bankrupt group.

Therefore, the posterior probability of bankruptcy knowing that the firm score ranges within the interval $i$ can be calculated using the Bayes’ theorem, such as:

$$p_{i} = \frac{\pi_{B} f_{i}^{B}}{\pi_{B} f_{i}^{B} + \pi_{NB} f_{i}^{NB}}$$

where $\pi_{B}$ and $\pi_{NB}$ are respectively the a priori probabilities of bankruptcy and non-bankruptcy of a firm within the next three years of exercise, hence: $\pi_{NB} = 1 - \pi_{B}$. The bankruptcy rates are computed by the French National Institute for Statistics and Economic Studies (INSEE). For the hospitality/ accommodation industry (NAF2: 55), the rate of bankruptcy is equal to 1.06%\(^{27}\) on a yearly basis or 3.18% if the rate remains constant in the future.

9.2. Risk classes and probabilities of bankruptcy of the score

Probabilities of bankruptcy (pi) are grouped in two years’ period, because of the small amount of bankruptcies available per year. This presents the advantage to make the results stronger while decreasing the temporal stability. Table 19 depicts the situation.

Table 19: Posterior probabilities of bankruptcy within three years horizon

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NB</td>
<td>B</td>
<td>Pi (%)</td>
<td>NB</td>
<td>B</td>
</tr>
<tr>
<td>2 ≤ z</td>
<td>18</td>
<td>1</td>
<td>0.18</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>0.4 ≤ z &lt; 2</td>
<td>21</td>
<td>6</td>
<td>0.93</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>-0.3 ≤ z &lt; 0.4</td>
<td>7</td>
<td>4</td>
<td>1.84</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>-0.8 ≤ z &lt; -0.3</td>
<td>2</td>
<td>3</td>
<td>4.70</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>z &lt; -0.8</td>
<td>1</td>
<td>35</td>
<td>53.48</td>
<td>1</td>
<td>20</td>
</tr>
</tbody>
</table>

NB: Non-bankrupt firms; B: Bankrupt firms; Pi: Posterior probability of bankruptcy within three years

For years 2008 and 2009 the horizons of failure are respectively two years and one year. Therefore, these years are not included in calculating the posterior probability retained for the score. The three years’ horizon posterior probability appears in the last column of Table 19. It is computed as the average of posteriors probabilities of periods 2002-2003, 2004-2005, and 2006-2007, which all have data for a three years’ horizon. The average is computed without weighing per number of firms present in each period. This method has the advantage to consider the temporal average of the posterior probabilities rather than being influenced too much by exceptional periods of bankruptcies, such as a financial crisis for example.

9.3. Uncertainty associated with probability of failure by risk class and its risk coefficient

Posterior probabilities (pi) may be a little uncertain depending on cyclical movements and sample biases; therefore, it is interesting to evaluate their degree of certainty. If per interval i, posteriors probabilities follow a normal distribution N(μ,σ^2), a confidence interval can be found such as:

\[
I_i(\alpha) = \left[ \mu_i - z_\alpha \frac{\sigma_i}{\sqrt{n-1}} ; \mu_i + z_\alpha \frac{\sigma_i}{\sqrt{n-1}} \right]
\]
where $z_\alpha$ is the upper $(1-\alpha)/2$ critical value for the standard normal distribution $N(0,1)$ and $n$ is the number of periods observed. Confidence intervals for $\alpha=0.1$ and $n=3$ are depicted in Table 20.

Table 20: Risk classes and their posterior probabilities on a three years’ horizon, with a confidence interval and a risk coefficient

<table>
<thead>
<tr>
<th>Intervals score</th>
<th>Risk classes</th>
<th>$\mu$ Pi (%)</th>
<th>$\sigma$ Pi (%)</th>
<th>Confidence interval at 0.1</th>
<th>Risk coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2 \leq z$</td>
<td>1</td>
<td>0.15</td>
<td>0.11</td>
<td>0.04 0.25</td>
<td>0.05</td>
</tr>
<tr>
<td>$0.4 \leq z &lt; 2$</td>
<td>2</td>
<td>1.22</td>
<td>0.84</td>
<td>0.42 2.02</td>
<td>0.38</td>
</tr>
<tr>
<td>$-0.3 \leq z &lt; 0.4$</td>
<td>3</td>
<td>3.08</td>
<td>1.37</td>
<td>1.78 4.38</td>
<td>0.97</td>
</tr>
<tr>
<td>$-0.8 \leq z &lt; -0.3$</td>
<td>4</td>
<td>5.97</td>
<td>4.18</td>
<td>2.00 9.94</td>
<td>1.88</td>
</tr>
<tr>
<td>$z &lt; -0.8$</td>
<td>5</td>
<td>36.20</td>
<td>15.71</td>
<td>21.28 51.11</td>
<td>11.38</td>
</tr>
</tbody>
</table>

Table 20 shows that the interval scores are regrouped in five homogenous risk classes, ranging from the safest (class 1) to the riskiest (class 5). Each class has a posterior probability ($p_i$) and a standard deviation ($\sigma$) with an explained interval of confidence. The last column of Table 20 presents the risk coefficient. It is evident that companies classified in class 5 are 11.38 times riskier than the reference bankruptcy rate observed on all companies file INSEE over a three years’ horizon (3.18%). At the opposite, companies in class 1 are 20 times less risky than the reference bankruptcy rate. Therefore, companies that are classified in class 1 and 2 constitute a favorable area. Companies classified in class 3 constitute a neutral class or a grey area. In this last class, non-bankrupt and bankrupt companies are present in a similar proportion, and their posterior probability of bankruptcy is consequently closer to the reference rate of bankruptcy from INSEE. There is some uncertainty remaining on the neutral zone because companies cannot be classified as bankrupt or non-bankrupt. Finally, companies in class 4 and 5 constitute an unfavorable zone.

In addition, the figure in Appendix J depicts the different classes of risks superimposed to the score distributions of the opposing two groups.
10. Analysis of the contributions

In this section are presented answers helping to determine the contributing variables to a predicted bankrupt and the benefit from it. To this end, different parts are presented, namely: contributing variables to a predicted bankrupt, the contributions’ meanings, the contributions of the industry, the contributions for firms, and the analysis of the scores through the period studied.

10.1. Contributing variables to a score or predicted outcome

10.1.1. First formulation of the score function – classification

The final function is composed of five ratios and appears in its discriminatory form as follows:

\[ z = -\alpha_1 x_1 - \alpha_2 x_2 + \alpha_3 x_3 - \alpha_4 x_4 + \alpha_5 x_5 + \beta \]

which produces:

\[ z = -2.47x_1 - 2.85x_2 + 0.84x_3 - 0.82x_4 + 0.75x_5 + 5.46 \]

where \((\alpha_1, \alpha_2, ..., \alpha_5)\) are the five coefficients vectors of the linear discriminant function, \((x_1, x_2, ..., x_5)\) are the vectors of the five ratios of a company, and \(\beta\) is the constant.

10.1.2. Second formulation of the score function – contributions

As already mentioned in the methodology section, it is more interesting to understand the different reasons that conduct to a certain score. This process is possible\(^{28}\), and the previous function can be rewritten in the practical form that follows:

\[ z = -\alpha_1 (x_1 - P_1) - \alpha_2 (x_1 - P_1) + \alpha_3 (x_3 + P_3) - \alpha_4 (x_4 - P_4) + \alpha_5 (x_5 - P_5) \]

which produces:

\[ z = -2.47(x_1 - 0.67) - 2.85(x_1 - 0.97) + 0.84(x_3 + 0.03) - 0.82(x_4 - 1.34) \]

\[ + 0.75(x_5 - 0.11) \]

where \((\alpha_1, \alpha_2, ..., \alpha_5)\) are the five coefficients vectors of the linear discriminant function, \((x_1, x_2, ..., x_3)\) are the vector of the five ratios of a company, and \((P_1, P_2, ..., P_5)\) are the vectors of the pivot values, which subtracted to vectors \((x_1, x_2, ..., x_5)\), furnish the five contributions of the ratios to the

\(^{28}\) See the methodology section for further detailed explanation
score. The pivot is calculated as such in a way that if a ratio is superior, the contribution of this ratio to the score is positive; in the opposite case, the contribution is negative.

10.2. Different contributions and their meanings

As depicted in Equations (10.1) and (10.2), the final model is composed of five elements contributing to its score ($z$). The contribution 1 is associated to the Current Liabilities’ ratio. When this last ratio increases, it means that more short-term debts are supported by the firm, resulting in a difficult situation; therefore, the contribution becomes negative to the total score. The contribution 2 is associated to Debt ratio – when this ratio increases, more debts are supported by the firm, describing a more difficult situation; therefore, the contribution becomes negative to the total score. The contribution 3 is associated to Return on Assets – when this ratio increases, returns are higher and describe (or present) a better situation; therefore, the contribution becomes positive to the total score. The contribution 5 is related to Cash-Flow to Short-Term Debt – when the ratio increases, more cash-flow is available – that is a good sign; therefore, the contribution becomes positive to the total score.

Contribution 4 is related to Asset Turnover ratio. Compared to the previous contributions, the interpretation of contribution 4 is more complex. This difficulty comes from the fact that either a low or a high ratio of asset turnover can be interpreted as a good sign that reflects a different firm’s strategy (Fairfield & Yohn, 2001). Firms with a higher asset turnover tend to have a lower profit margin, and on the contrary, firms with a lower asset turnover tend to have a higher profit margin. In practice, the ratio of asset turnover and its contribution (see in Equation (10.2) – contribution 4) considers both situation of the operating strategy (high profit margin vs. low profit margin). This is possible because of the relatively large pivot value ($P_4$) that is set in the Equation (10.2). Therefore, depending on the value of the asset turnover ratio to the pivot value, contributions are positive or negative to the final score. In addition, readers may refer to Altman (1968), who pinpointed this interpretation difficulty.

Consequently, with the help of contributions, the final score provides synthetized information of the good health of a firm’s situation. It helps in pinpointing the strong aspects (positive contributions) from the weak ones (negative contributions). By using this interesting method, it becomes possible to analyze different situations of the firm’s life.
In the following two parts, different situations are approached. These different situations are (i) the analysis of the industry performances as benchmark, and (ii) the analysis of specific firm’s scenarios, such as: non-bankrupt, turnaround and bankrupt situations.

10.3. Contributions of the sector

The model once segmented in its different contributions can be of a great interest. One of them is to analyze the different contribution results of the whole industry by quantiles. Figure A in Appendix K depicts such a situation. In this figure, the five different contributions are depicted as well as the final combination of the scores. All these contributions and scores are detailed by Q1, Q2, and Q3. The first quartile (Q1) represents the value at 25% of the sample studied; the second or median quartile (Q2) represents the value at 50% of the sample studied; and the third quartile (Q3) represents the value at 75% of the sample studied.

At first, it can be noticed that the scores for the hospitality/ accommodation industry (NAF2: 55), remain constant within the period studied. In detail, it can be seen that the scores declined slightly around years 2006 and 2007. This decline may reflect the overall impact of the recent financial crisis on this sector. Therefore, in this period, scores are lower, meaning that firms represent a higher risk than in normal period.

The reasons of a certain score can be explained from the five contributions detailed. For example, for the lower industrial scores in 2006, this situation can be explained by lower contribution 1 (heavier short-term debts) and contribution 5 (lesser cash-flows), while other contributions remain constant. In addition, further analyses of the score and its contributions can be conducted on practitioners’ needs using the same principle.

10.4. Contributions for firms

Using the same principle, specific firms performances are analyzed and benchmarked against other competitors or industries quantiles. In the following paragraphs, three specific situations encountered by a firm are presented, such as: (i) behavior of a non-bankrupt firm, (ii) behavior of a turnaround firm, and (iii) behavior of a bankrupt firm.
10.4.1. Case of a non-bankrupt firm

From figure B in Appendix K, it can be noticed that contribution 1 and contribution 4 went down during the financial crisis of 2007. Contribution 2 went up in the meantime, reflecting a swap between long to short-term debt in the firm’s strategy (possibly due to some bank restrictions on loans during the crisis). Therefore, the score of the firm went down mainly because of contribution 1 and 4 that went down during the financial crisis of 2007.

After the situation was cleared on these last contributions (C1 and C4), the firm recovered its previous score level.

10.4.2. Case of a turnaround firm

From figure C in Appendix K, it can be noticed (from the window depicting the scores in the bottom right corner), that results for this firm are not good until year 2007, when some changes appear in its management. In particular, this improvement in the score result is due to an improvement of contribution 1, where the weight of the short-term debt decreased, positioning the firm in a better situation. In addition, contribution 5 increased, which means that the firm received more cash-flow. Finally, more situations could be described with the same process on practitioners’ needs.

10.4.3. Case of a bankrupt firm

Finally, one last situation could be encountered in the business life of a firm – this is the process leading to a bankruptcy. With the help of the contributions’ analysis, it becomes possible for a specific firm to analyze the potential reason for a bankrupt, and if possible to remedy it.

From figure D in Appendix K, it can be noticed that the firm’s score was low for several years, up to the year of its bankruptcy – 2009. For this specific firm, the bankrupt resulted in a combination of poor performances on all ratios. Contribution 1 depicts the weight of the short-term debt, which is heavy since 2001. Contribution 2, 3, and 5 show worse outcomes, resulting in a poorer situation. On the other side, contribution 4 appeared to progress a little bit, but this is in fact due to the conception and complexity of this ratio, as previously mentioned.

---

29 In this last situation (C4), the contribution only increased because the firm had probably sold many assets or at least they were depreciated, hence the fictive increase on this contribution.
10.5. Analysis of the scores through the period studied

In addition, if placed in the context of the whole period studied, results from different firms can be compared within and between each other for several benefits. The fact to highlight in different colors the different classes of risk, as mentioned in the previous section, can help in analyzing a company’s situation.

Table 21 below depicts the scores of different profiles of firms and their corresponding degree of risk. As previously mentioned, the five different classes of risk are represented in the table below. They are colored from the darkest and riskiest (No.5) to the clearest and safest (No. 1). Finally, when firms are bankrupt, they are marked with “B”.

<table>
<thead>
<tr>
<th>Firms</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.574</td>
<td>2.176</td>
<td>2.013</td>
<td>1.780</td>
<td>1.593</td>
<td>1.346</td>
<td>1.218</td>
<td>1.253</td>
<td>1.895</td>
<td>2.189</td>
</tr>
<tr>
<td>B</td>
<td>-0.192</td>
<td>-0.194</td>
<td>1.305</td>
<td>2.101</td>
<td>1.252</td>
<td>2.220</td>
<td>1.718</td>
<td>1.008</td>
<td>1.519</td>
<td>2.893</td>
</tr>
<tr>
<td>C</td>
<td>-0.144</td>
<td>1.310</td>
<td>1.716</td>
<td>1.399</td>
<td>2.041</td>
<td>1.025</td>
<td>1.612</td>
<td>0.885</td>
<td>-0.301</td>
<td>-1.054</td>
</tr>
<tr>
<td>D</td>
<td>2.115</td>
<td>2.342</td>
<td>2.139</td>
<td>1.843</td>
<td>1.571</td>
<td>2.280</td>
<td>2.412</td>
<td>5.563</td>
<td>6.076</td>
<td>3.082</td>
</tr>
<tr>
<td>E</td>
<td>0.074</td>
<td>0.287</td>
<td>-0.566</td>
<td>0.119</td>
<td>-0.101</td>
<td>-0.906</td>
<td>-0.178</td>
<td>1.327</td>
<td>1.062</td>
<td>0.695</td>
</tr>
<tr>
<td>F</td>
<td>0.842</td>
<td>2.110</td>
<td>2.765</td>
<td>2.046</td>
<td>2.316</td>
<td>2.416</td>
<td>1.840</td>
<td>2.371</td>
<td>1.789</td>
<td>2.870</td>
</tr>
<tr>
<td>G</td>
<td>6.096</td>
<td>1.580</td>
<td>1.547</td>
<td>1.540</td>
<td>1.943</td>
<td>1.926</td>
<td>1.832</td>
<td>1.470</td>
<td>2.056</td>
<td>1.691</td>
</tr>
<tr>
<td>J</td>
<td>-0.298</td>
<td>0.330</td>
<td>-0.594</td>
<td>-0.127</td>
<td>-0.824</td>
<td>-0.220</td>
<td>-0.288</td>
<td>-1.326</td>
<td>-1.689</td>
<td>B</td>
</tr>
<tr>
<td>K</td>
<td>-0.575</td>
<td>-2.018</td>
<td>-1.645</td>
<td>-1.630</td>
<td>-1.142</td>
<td>0.567</td>
<td>-0.682</td>
<td>-0.181</td>
<td>-0.364</td>
<td>B</td>
</tr>
</tbody>
</table>

From Table 21, it can be noticed instantaneously how well a firm performs. This table can help in rapidly determining if a firm is worth to invest in, such as in the turnaround case (i.e. firm E before 2007, firm C from 2010, if available for sale). It also helps managers to position themselves toward competitors in the context of risk categories of default. It is also useful for credit managers that need to facilitate credit to a company, and so forth. Finally, it is noticeable that firms A, E, and J are respectively firms that are detailed in (i) the analysis of a non-bankrupt firm, (ii) the analysis of a turnaround firm, and (iii) the analysis of a bankrupt firm.

The next section presents an overall discussion of the study accomplished.
11. **Discussion**

The following section discusses in a more general perspective observations from the study realized. These observations are organized into the following four parts. First, remarks are associated to the modeling technique used. Second, observations are discussed from the preparation of the discriminatory function. Third, problems results from the model application. Finally, some lateral and other issues are discussed.

11.1. **Modeling technique used**

The choice of the modeling technique is determined by the practitioners’ needs and emphasis. As previously mentioned in the literature section, there is no technique that is superior to the others. Each one possesses its pros and cons, which is probably the reason why combining different techniques (i.e. soft computing) has become popular in the recent years. However, many researchers have noticed that the simplest models often outperform the most complicated ones (Bardos, 2001). The linear discriminant analysis (LDA) technique is used in this study, because of its likely high predictive ability, robustness to future applications, and interpretability.

The frequent problem concerning the LDA technique is that it relies on too many assumptions. Specifically, three assumptions make it difficult to implement. They are: the assumption of multivariate normality, the assumption of homoscedasticity, and the assumption of linearity.

To begin, the multivariate normality assumption is empirically difficult to implement as financial data are not totally adequate to this assumption (Balcaen et al., 2006). Therefore, it is intended empirically to approach this assumption by approximating the univariate normality. However, as mentioned in the robustness analysis section, univariate normality is not a sufficient condition for multivariate normality, and even if empirically the model is robust to the violation of this assumption, the problem remains from a theoretical perspective.

Another problem concerns the assumption of equal dispersion matrices that is frequently violated. Even if the model appears to be robust empirically to the loss of this assumption, there is again a problem from a theoretical perspective. However, results from testing this assumption may be biased by the violation of the previous assumption on multivariate normality. In attempt to make
the model fit closer to this assumption, practitioners can use a quadratic discriminant analysis (QDA) model that does not rely on the assumption of equal dispersion matrices. However, researchers demonstrate that mostly of the time, the best results are achieved by doing LDA (Karels & Prakash, 1987). The reason for this superiority is that the QDA with one assumption dropped to the LDA, becomes very sensitive to the assumption of multivariate normality. Thus, results are too relying on the latter assumption that is frequently violated.

The final remark concerns the assumption of linearity, which limits the study’s model to variables that demonstrate a linear relationship to bankruptcy status. Other variables that are discriminatory between opposing groups (bankrupt and non-bankrupt firms), cannot be considered with this model. Therefore, the model’s capacities to predict accurate rate of good classification are limited to this extent.

11.2. Preparation of the discriminatory function

Another category of problems concerns the preparation of the model function, which covers the type of data used, the data reprocessing, the selection of the sample, and the subjectivity of the model.

To begin, other sources of information than financial statement data could be used in a similar study. The choice of the dataset is subjective to the practitioners, as well as result from the model and study developed. Other sources of information include: macroeconomic data (i.e. inflation, internal or external demand, GDP, exchange rates), stock exchange data and quotation, and qualitative data. They all have pros and cons. Financial statement data are widely available as the most of the firms are legally obliged to disclose financial statement data to regulatory authorities. The drawback of using financial statement data is that they are cross sectional information, which does not consider inputs over time. Consequently, it is more valuable to compare the results of a given company related to its previous results.

Another problem concerns the data reprocessing that may disturb the results of the technique implemented. As outputs of the LDA method are dependent on averages and covariance matrices, severe modifications on these inputs could bias the results’ output. As previously mentioned in the data reprocessing section, the best thing to do is to allow for data reprocessing for variables that really need it (Bardos, 2001).
Further remark concerns whether or not the sample constituted should be selected on a random basis from the whole population. In this study, data for non-bankrupt firms are selected on a stratified random basis without replacement to match the number of bankrupt firms. According to Bardos (2001), this matching should enhance the model’s robustness against violation of the two assumptions of multivariate normality and homoscedasticity. In addition, bankrupt firms being rare, it appears preferable to use the most of the data available in order to make the study as robust as possible. On the other hand, other researchers suggest that the whole population should be reflected on a random basis (Balcaen et al., 2006).

Finally, the idea of working with a quantitative model and data increases the objectivity of this study. However, there is some subjectivity associated in the development of the model. Specifically, the subjectivity concerns the selection and association of the discriminant ratios in the final model. As mentioned in the development of the model section, different techniques are used to develop the model such as the stepwise analysis, decision trees, financial analysis and so forth. Thus, selection of discriminatory variables into the final model depends on the subjective emphasis of the practitioner towards the results of these techniques. There is no claim for optimality, even if the model selected aims to be the best among the ones available (Altman, 2006).

11.3. Model application

Different problems result from the model’s application. These observations are discussed into the following four parts: (i) measurement of the performance, (ii) ignored costs of errors, (iii) probabilities of bankruptcy, and (iv) model’s interpretation.

Depending on the measured performance, different results can be interpreted. In this study, the overall good rate of classification is chosen to measure the performance of the model. It reflects the fact that both bankrupt and non-bankrupt firms are considered equally important in the bankruptcy prediction. However, this importance is subjective and different goals could be set if needed by practitioners. For example, with proper adjustments, investors could determine numbers of firms’ worth to invest in30.

---

30 Refer to the empirical results and contributions sections for further and detailed examples
Another point concerns the costs of type I and type II errors that, if ignored, lead to relative high type I errors (Balcaen et al., 2006). However, as demonstrated in the empirical results’ section, it is possible to overcome the over representation of type I errors by changing the cut-off value. In this study, this is achieved by nonlinear programming for increased performance.

Further observation concerns the probabilities of bankruptcy that are relatively subjective to the practitioner. Intervals’ score can be adjusted to different values, which influences corresponding probabilities of bankruptcy (p_i). In addition, in the calculation of the posterior probability of bankruptcy through the Bayes’ theorem, it is assumed that the rate of bankruptcy of the industry remains constant in the future, which might not be the case. Equally important, corresponding confidence intervals assume that posterior probabilities follow a normal distribution, which again might not be the case.

Finally, even if the model interpretation has been extended, some difficulties persist in the model’s interpretation and analysis of the contribution. Specifically, three difficulties can appear. First, multicollinearity among variables can cause severe misleading results. Second, some ratios may be complex to interpret such as the Asset Turnover ratio mentioned in the previous section. Third, conflicting variables with the linearity assumption may disturb the interpretation. However, once these difficulties are well understood, the model interpretation becomes very interesting and proposes several possibilities such as the possibility to compare on table different companies over time.

11.4. Later and other issues

In addition, other problems concern: (i) definition of bankruptcy, (ii) evolution of economic environment, (iii) limitations of focused model, and (iv) data not including bankrupt’s symptoms.

To begin, there are some issues regarding the definition of bankruptcy that may be different from one study to another. Three problems can be discussed (Balcaen et al., 2006). At first, the definition of bankruptcy is relatively subjective. Second, the bankruptcy process is not a perfect dichotomy. Third, there are some specific problems using the bankruptcy term to discriminate groups such as: (i) models may be biased by variables corresponding to bankruptcy (i.e. solvency

---

31 Ohlson (1980) pointed out that one of the drawback of the LDA was its poor interpretation, which has been solved ever since with the second formulation of the score function.
or liquidity); (ii) models might not consider other bankrupt situations like: voluntary bankrupt as a strategy, merger, liquidation, and absorption among others.

Another problem concerns the fact that companies evolve and change over time, which might mislead results. Specifically, two concerns can be discussed. First, the model considers solely observations from one single period. Thus, a firm could be drastically classified as in difficulty whereas in reality, its situation could depict a small temporary difficulty or a change in strategy. Consequently, it becomes valuable to analyze the situation of a firm over time and create a dynamic. Second, the data of the firms may change over time. Thus data instability and non-stationarity lead to misleading results, as coefficients in the discriminatory function are remaining the same. However, different possibilities exist to adjust the model to this data instability (Planès, 2004).

Further, the fact to focus the model on a specific industry is debatable. There are different pros and cons in focusing a model. The advantages are to tailor the study technique to a specific need or goal targeted. Another advantage is that based on this specificity results might be higher. On the other hand, an evident drawback is that the model cannot be directly implemented on another industry population. Another drawback might be that the model is too specific to the sample studied and might not be accurate with newer data. However, tests on a holdout sample should assure that the model is applicable to newer data.

Finally, problems concerning the symptoms describing the phenomenon of bankruptcy have to be included in the data information for an efficient model use. However, two limitations appear. First, if the firm does not possess financial statement information (i.e. micro and start-up firms). Second, firms in difficulty may mislead or trick their financial statement information (i.e. Enron, 2001). Thus, without or with wrong information the developed model may not forecast accurately the situation. In order to palliate these possibilities, additional approaches covering other characteristics of a firms’ situation might be beneficial (i.e. by relying on the expert knowledge or qualitative information). The point with a focused model is to develop on its specific domain, a very robust model, rather than the most general model that would underperform in the most of the situations.

The next section concludes this study.
12. Conclusion and future direction

This last section presents conclusions on the study and future directions for potential follow-ups. The initial motivation of this study was to understand the phenomenon leading to bankruptcy and to find elements allowing to benefit from it. Consequently, the research questions that defined this study were: (i) How to predict bankruptcies on a specific industry? (ii) How to attribute probabilities of bankruptcy and classes of risk to these predictions? (iii) How to determine the contributing variables to a predicted bankrupt and to benefit from it?

In order to answer the first question raised, five sections were developed: (i) literature review, (ii) theoretical model, (iii) data, (iv) model development, and (v) empirical results. At first, an exhaustive literature review was presented to determine what technique could suit best the topic studied. This review ranged from historical insights and first major studies to the latest researches up to this date. From this review, the linear discriminant analysis (LDA) was chosen as the most appropriate to predict bankruptcy on a specific industry. Most importantly, this statistical technique as a benchmark method in bankruptcy prediction was likely to procure a good rate of classification, a robustness to economic changes, as well as a strong interpretation ability, needed to answer the remaining research questions. Further, in order to implement well this method, a detailed presentation of the theoretical methodology was presented.

Then, data were gathered and prepared. For this matter, the data section presented the specificities of the data, such as: definition of bankruptcy, limitations and applications of the model to the study, and some descriptive statistics. In addition, different data reprocessing was applied. This reprocessing ranged from the softest (winsorization of extreme values) to the hardest (both numerator and denominator could change their signs).

Then, different models were developed in two successive steps: (i) at first, ratios were selected according to their ability to discriminate opposing groups (bankrupt and non-bankrupt firms). To this end, different steps and tests were conducted. First, histograms and Kolmogorov tests were used followed by quantiles and median tests, and finally, tests on correlations: Spearman and Pearson tests. (ii) As a second step, combinations of discriminatory ratios were assembled in different models using various techniques (i.e. stepwise analysis).
These combined models with different characteristics were tested on the training data. One final model was selected and improved by nonlinear programming in order to balance in equal proportion the classification of bankrupt and non-bankrupt firms. Then, the improved model was finally tested on the holdout sample that accounted for almost 50% of the data available. The rate of good classification was equal to 86.36% and classification results of firms were balanced between two opposing groups. In addition, various tests on the robustness of the developed model were performed. The majority of these tests demonstrated the good performance of the final model. This paved the way for answering the first research question.

In order to further explore the benefits from the model developed and to answer the last two research questions, it was necessary to refine the binary classification with some additional outputs, such as: (i) the analysis of probabilities of bankrupt and classes of risk, and (ii) the analysis of the variables contributing to this final score. (i) Regarding the first analysis, the final score was segmented into five different classes with probabilities of bankruptcy. These classes of risk depicted different situations. They classified firms from the most to the least risky one. Then, corresponding confidence intervals were associated to refine this analysis. (ii) The final analysis consisted of understanding the outputs of the score classification. Causes leading to a certain outcome over time were analyzed. Finally, discussions on the study performed were presented.

In summary, this study contributed to the literature review in this domain; developed and used a model for a specific industry or need; used nonlinear programming to optimize the cut-off value, and offered some additional insight and perspective to analyze the benefits of using this model (i.e. posterior probabilities of bankruptcy, classes of risk, contributions to the score over time, and discussion sections). In addition, this study, as some other studies, performed some testings on general model robustness (i.e. check of major assumptions and testing of the significance of the variables through logistic regression).

Regarding potential follow-ups, it would be possible to complement this study with additional data inputs, either quantitative or qualitative. Results from this study could be compared to the results of other predicting techniques, such as support vector machine, soft computing, and dynamic modeling. In order to perform a fair comparison, same data among different techniques should be used. Finally, this study could be applied to any other industry or specific need, as long as the same methodology is respected for the benefit of all interested parties.
References


Roy, B., Benayoun, R., & Sussman, B. (1966). ELECTRE : Une m’ethode pour guider le choix en pr’esence de points de vue multiples. Note De Travail 49, SEMA-METRA International, Direction Scientifique,

Roy, B. (1968). Classement et choix en présence de points de vue multiples (la méthode ELECTRE). Revue d’Informatique Et De Recherche Opérationnelle, 8, 57-75.


Stili, D. (2002). La détection précoce des défaillances d'entreprises dans le secteur de la construction, score BDFB. Banque De France, Direction Des Entreprises, Cahiers Études Et Recherches De l'Observatoire Des Entreprises,


Appendix A – Initially selected ratios for the study

A. Liquidity

1. Current ratio
   \[
   \frac{\text{current assets}}{\text{current liabilities}}
   \]

2. Quick ratio or acid test ratio
   \[
   \frac{\text{cash + marketable securities + receivables}}{\text{current liabilities}}
   \]

3. Cash Ratio
   \[
   \frac{\text{cash + marketable securities}}{\text{current liabilities}}
   \]

4. Working Capital to Total Assets Ratio
   \[
   \frac{\text{Working Capital}}{\text{Total Assets}}
   \]

5. Inventory to Working Capital Ratio
   \[
   \frac{\text{Inventory}}{\text{Accounts receivable + Inventory - Accounts payable}}
   \]

6. Working Capital to Short-term Debt Ratio
   \[
   \frac{\text{Cash + Accounts receivable + Inventory - Accounts payable}}{\text{Short-term Debt}}
   \]

7. Working Capital to Long-term Debt Ratio
   \[
   \frac{\text{Cash + Accounts receivable + Inventory - Accounts payable}}{\text{Long-term Debt}}
   \]

8. Working Capital to Total Debt Ratio
   \[
   \frac{\text{Cash + Accounts receivable + Inventory - Accounts payable}}{\text{Total Debt}}
   \]

9. Sales to Current Assets Ratio
   \[
   \frac{\text{Sales}}{\text{Current Assets}}
   \]

10. Working Capital Productivity
    \[
    \frac{\text{Sales}}{\text{Working Capital}}
    \]

11. Current Liabilities Ratio
    \[
    \frac{\text{Current Liabilities}}{\text{Total Liabilities}}
    \]

12. Non-Current Assets to Non-Current Liabilities Ratio
    \[
    \frac{\text{Non-current Assets}}{\text{Non-current Liabilities}}
    \]

13. Short-term Debt to Long-term Debt Ratio
    \[
    \frac{\text{Total short-term Debt}}{\text{Total Long-term Debt}}
    \]

14. Working Capital to Cash Expenses
    \[
    \frac{\text{Working Capital}}{\text{Cash Expenses}}
    \]
### B. Leverage

15. Debt Ratio
   \[ \frac{Total Liabilities}{Total Assets} \]

16. Long-term Debt to Equity Ratio
   \[ \frac{Long-term Debt}{Total Assets} \]

17. Capitalization Ratio
   \[ \frac{Long-term Debt}{Long-term Debt + Owners' Equity} \]

18. Cash Coverage Ratio
   \[ \frac{EBIT + depreciation - interests payments}{Net income + Interest Expense + Income Tax Expense} \]

19. Debt to Equity Ratio
   \[ \frac{Long-term Debt + Value of Leases}{Shareholders' Equity} \]

20. Times Interest Earned
   \[ \frac{Net Income + Interest Expense + Income Tax Expense}{Interest Expense} \]

21. Long-term Liabilities to Total Assets
   \[ \frac{Long-term Liabilities}{Total Assets} \]

22. Funded Capital Ratio
   \[ \frac{Stockholders' Equity + Long-term Debt}{Fixed Assets} \]

23. Retained Earnings to Stockholders’ Equity
   \[ \frac{Retained Earnings}{Total Stockholders' Equity} \]

24. Current Liabilities to Equity
   \[ \frac{Current Liabilities}{Equity} \]

25. Times Preferred Dividend Earned
   \[ \frac{Net Income}{Preferred Dividend} \]

### C. Solvability

26. Interest Coverage Ratio
   \[ \frac{EBIT}{Interest Expense} \]

27. Long-term Debt to Net Working Capital
   \[ \frac{Current Assets - Current Liabilities}{Net operating Income} \]

28. Debt service coverage ratio
   \[ \frac{Net income + Interest Expense + Income Tax Expense}{Total Debt Service} \]
## D. Profitability

29. EBT to Revenues Ratio  
   \[
   \frac{EBT}{Revenues}\n   \]

30. Net Profit Margin  
   \[
   \frac{Net Income}{Revenues}\n   \]

31. Operating Income Margin  
   \[
   \frac{Operating Income}{Net Sales}\n   \]

32. Operating Profit Margin  
   \[
   \frac{Net Income + Interest}{Sales}\n   \]

33. Return on Assets  
   \[
   \frac{Net Income}{Total Assets}\n   \]

34. Return on Assets – Advanced version  
   \[
   \frac{Net Income + Interest Expense, Net of Tax}{Total Assets}\n   \]

35. Return on Equity  
   \[
   \frac{Shareholders' Equity}{Net Income}\n   \]

36. Return on Capital Employed  
   \[
   \frac{EBIT}{Equity}\n   \]

37. Basic Earnings Power Ratio  
   \[
   \frac{EBIT}{Total Assets}\n   \]

38. Return on Capital  
   \[
   \frac{EBIT (1-Tax Rate)}{Invested Capital}\n   \]

39. Net Gearing  
   \[
   \frac{Net Debt}{Equity}\n   \]

40. Financial Leverage Index  
   \[
   \frac{Return on Equity}{Return on Assets}\n   \]

41. Gross Profit Margin  
   \[
   \frac{Gross Profit}{Net Sales}\n   \]

42. Dividend Payout Ratio  
   \[
   \frac{Dividend per share}{Earnings per share}\n   \]

43. Dividend Yield Ratio  
   \[
   \frac{Dividend per share}{Market price per share}\n   \]

44. Net Worth  
   \[
   \frac{Total Assets - Total Liabilities - Preferred stock dividends}{Total Outstanding Common Shares}\n   \]
E. Operating Performance Measurement

45. Sales to Operating Income Ratio
   \[ \frac{\text{Operating Income}}{\text{Net Sales-Investment Income}} \]
   \[ \frac{\text{Gross Margin-Sales expenses}}{\text{Gross Sales}} \]
   \[ \frac{\text{Sales-Variable expenses}}{\text{Operating income}} \]

46. Sales Margin

47. Operating Leverage Ratio

F. Efficiency

48. Asset Turnover
   \[ \frac{\text{Net Sales}}{\text{Total Assets}} \]
   \[ \frac{\text{Net Sales}}{\text{Net Fixed Assets}} \]
   \[ \frac{\text{Net Credit Sales}}{\text{Accounts Receivable}} \]

49. Fixed Asset Turnover

50. Receivables Turnover Ratio
   \[ \frac{\text{Sales}}{\text{Average Accounts Receivable}} \]
   \[ \frac{\text{Sales}}{\text{Net Sales}} \]
   \[ \frac{\text{Accounts Receivable}}{\text{Net Sales}} \]

51. Receivables Turnover in Days
   \[ \left( \frac{\text{Accounts Receivable}}{\text{Net Sales}} \right) \times 365 \text{ or } 360 \text{ days} \]

52. Inventory to Sales Ratio

53. Payables Turnover Ratio
   \[ \frac{\text{Purchases}}{\text{Average Accounts Payable}} \]
   \[ \frac{\text{Average Accounts Payable}}{\text{Purchases}} \]
   \[ \frac{\text{Cost of Goods Sold}}{\text{Average Inventory}} \]

54. Payables Turnover in Days
   \[ \left( \frac{\text{Average Accounts Payable}}{\text{Purchases}} \right) \times 365 \text{ or } 360 \text{ days} \]

55. Inventory Turnover Ratio

56. Inventory Turnover in Days
   \[ \left( \frac{\text{Average Inventory}}{\text{Cost of Goods Sold}} \right) \times 365 \text{ or } 360 \text{ days} \]

G. Cash flow ratio

57. Cash Flow to Short-term Debt Ratio
   \[ \frac{\text{Net Income} + \text{Noncash expenses-Noncash sales}}{\text{Short-term debt}} \]

58. Cash Flow to Long-term Debt Ratio
   \[ \frac{\text{Net Income} + \text{Noncash expenses-Noncash sales}}{\text{Long-term debt}} \]
59. Cash Flow to Total Debt Ratio  
\[
\frac{\text{Net Income} + \text{Noncash expenses-Noncash sales}}{\text{Total debt}}
\]

60. Cash Flow Return on sales  
\[
\frac{\text{Net Income} + \text{Noncash expenses-Noncash sales}}{\text{Total sales}}
\]

61. Cash Flow Return on Assets  
\[
\frac{\text{Net Income} + \text{Non cash expenses-Non cash sales}}{\text{Total Assets}}
\]

62. Cash Flow from Operations to Current Liabilities Ratio  
\[
\frac{\text{Net Cash Provided by Operating Activities}}{\text{Average Current Liabilities}}
\]

63. Cash Flow to Net Worth  
\[
\frac{\text{Cash Flow}}{\text{Net Worth}}
\]

64. Cash to Current Assets Ratio  
\[
\frac{\text{Cash} + \text{Short-term marketable securities}}{\text{Current Assets}}
\]

65. Cash to Working Capital Ratio  
\[
\frac{\text{Cash} + \text{Short-term marketable securities}}{\text{Current Assets-Current Liabilities}}
\]

H. Asset Utilization Measurements

66. Sales to Working Capital Ratio  
\[
\frac{\text{Net Sales}}{\text{Accounts Receivable + Inventory-Accounts Payable}}
\]

67. Sales to Equity Ratio  
\[
\frac{\text{Net Sales}}{\text{Total Equity}}
\]

68. Depreciation to Fixed Assets Ratio  
\[
\frac{\text{Depreciation}}{\text{Total Fixed Assets}}
\]

69. Interest Expense to Debt Ratio  
\[
\frac{\text{Interest Expense}}{\text{Short-term Debt + Long-term Debt}}
\]

70. Investment Turnover  
\[
\frac{\text{Sales}}{\text{Stockholders' Equity + Long-term Liabilities}}
\]

71. Tax Rate Percentage  
\[
\frac{\text{Income tax expense}}{\text{EBT}}
\]

72. Sales to Administrative Expenses Ratio  
\[
\frac{\text{Net Sales}}{\text{Total general and administrative expenses}}
\]

73. Break-Even Point  
\[
\frac{\text{Total Operating Expenses}}{\text{Average Gross Margin in Percentage}}
\]

74. Margin of Safety  
\[
\frac{\text{Current Sales Level - Break-Even Point}}{\text{Current Sales Level}}
\]

Note: numbers that are preceded by “/” represent ratios that could not be computed because of insufficient data information. Therefore, these ratios could not be considered further in the study.
Appendix B – Shortlisted ratios for the LDA

A. Liquidity
1. Current ratio
2. Quick ratio or acid test ratio
3. Cash ratio
4. Working Capital to Total Assets ratio
5. Working Capital to Short-term Debt ratio
6. Working Capital to Total Debt ratio
7. Sales to Current Assets ratio
8. Current Liabilities Ratio
9. Operating Profit Margin Ratio
10. Return on Assets
11. Return on Equity
12. Return on Capital Employed

B. Leverage
9. Debt Ratio
10. Long-term Debt to Equity Ratio
11. Cash Coverage Ratio
12. Funded Capital Ratio

C. Profitability
13. EBT to Revenues Ratio
14. Net Profit Margin Ratio
15. Operating Income Margin Ratio
16. Operating Profit Margin Ratio
17. Return on Assets
18. Return on Equity
19. Return on Capital Employed

D. Efficiency
20. Asset Turnover Ratio
21. Inventory to Sales Ratio
22. Cash flow to Short-term Debt Ratio
23. Cash flow to Total Debt Ratio
24. Cash flow Return on Assets Ratio

E. Cash flow ratio
25. Interest Expense to Debt Ratio
26. Tax Rate Percentage Ratio

F. Asset Utilization Measurements
Appendix C – BvD (Orbis) definition of a company’s status

A. Inactive

According to the Bureau van Dijk, Inactive firms are decomposed in seven sub-categories in the Orbis database, such as:

1. Bankruptcy: bankruptcy is a legally declared inability of a company to pays its creditors. The company no longer exists because it has ceased its activities since it is in the process of bankruptcy.

2. Dissolved: the company no longer exists as a legal entity, but the reason for this is not specified.

3. Dissolved (merger): the company no longer exists as a legal entity because the company has been included in a merger.

4. Dissolved (demerger): the company no longer exists as a legal entity, the reason for this is a demerger - the company has been "split".

5. In liquidation: the company no longer exists because it has ceased its activities, since it is in the process of liquidation.

6. Inactive (branch)

7. Inactive (no precision): the company is no longer active and the precise reason for inactivity is unknown.

Only firms within the bankruptcy status as stated in point 1 were selected in this study.

B. Active

On the other side, Active companies are decomposed in five categories such as:

1. Active: the company is active.

2. Active (default of payment): The term “default” should be distinguished from the terms “insolvency” and “bankruptcy”. "Default" essentially means a debtor has not paid a debt. "Insolvency “is a legal term meaning that a debtor is unable to pay his debts. "Bankruptcy" is a legal finding that imposes court supervision over the financial affairs of those who are insolvent or in default.

3. Active (dormant)

4. Active (receivership): the company remains active, though is in administration or receivership or under a scheme of arrangement (US - Chapter 11). During this period, the company is usually placed under the protection of a law and continues operating and repaying creditors and tries to reorganize and return to normal operating. At the end, the company will either return to normal operating (the default of payment was thus temporary), or will be reorganized (parts of its activity can be restructured or sold) or will be liquidated.

5. Active (branch)

Only firms within the active status as stated in point 1 were selected in this study.

32 Source: BvD (Orbis) – Company status definitions, User guide

33 Source: BvD (Orbis) – Company status definitions, User guide
Appendix D – Ratio reprocessing

Below are presented different types of reprocessing for the 26 ratios that were found discriminant in the study.

A. Reprocessing – type 1a

The reprocessing as depicted in Table 22 was applied to: current ratio, quick ratio, cash ratio, cash coverage ratio, cash flow to short term debt, cash flow to total debt, cash flow return on assets ratio, and interest expense to debt ratio.

<table>
<thead>
<tr>
<th>Cash-flow</th>
<th>Short-term debt</th>
<th>0</th>
<th>+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inferior boundary</td>
<td>Inferior boundary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superior boundary</td>
<td>Potential winsorization within limits [inf., sup]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B. Reprocessing – type 1b

The reprocessing as depicted in Table 23 was applied to: sales to current assets ratio, funded capital ratio, return on capital employed, and inventory to sales ratio (this later having approximately 30% of retreatments).

<table>
<thead>
<tr>
<th>Sales</th>
<th>Current Assets</th>
<th>0</th>
<th>+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inferior boundary</td>
<td>Inferior boundary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>Potential winsorization within limits [inf., sup]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
C. Reprocessing – type 1c

The reprocessing as depicted in Table 25 was applied to the following ratios: EBT to Revenues ratio, net profit margin ratio, operating income margin ratio, and operating profit margin ratio.

Table 24: Example of retreatment for the net profit margin ratio

<table>
<thead>
<tr>
<th>Net income</th>
<th>Revenues</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0 or = 0</td>
<td>Inferior boundary</td>
</tr>
<tr>
<td>+</td>
<td>Potential winsorization within limits [inf., sup]</td>
</tr>
</tbody>
</table>

D. Reprocessing – type 2a

This reprocessing as depicted in Table 25 was applied to the return on equity and to the return on capital employed ratios.

Table 25: Example of retreatment for the return on equity

<table>
<thead>
<tr>
<th>Net Income</th>
<th>Shareholders’ Equity</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0</td>
<td>Potential winsorization within limits [inf., sup]</td>
</tr>
<tr>
<td>+</td>
<td>Potential winsorization within limits [inf., sup]</td>
</tr>
</tbody>
</table>
E. Reprocessing – type 2b

This reprocessing as presented in Table 26 was solely applied to the tax rate percentage ratio.

<table>
<thead>
<tr>
<th>Table 26: Example of retreatment for the tax rate percentage ratio</th>
</tr>
</thead>
</table>
| ![Table](image)

F. Ratios that had no reprocessing

Eight ratios did not need to be reprocessed (except of the winsorization). They are: working capital to total assets ratio, working capital to short term debt ratio, working capital to total debt ratio, current liabilities ratio, debt ratio, long term debt to equity ratio, return on assets, and asset turnover ratio.
Appendix E – SAS coding

PROC UNIVARIATE DATA = WORK.SORTTempTableSorted;
   VAR "Debt Ratio"n;
   CLASS Bankruptcy;
   HISTOGRAM / CFILL=GRAY CAXES=BLACK WAXIS=1 CBARLINE=BLACK
               CFILL=BLUE PFILL=SOLID MIDPOINTS=-1 TO 10 BY 0.2;
/* End of task code. */
RUN; QUIT;
Appendix F – Data selection

A. Histograms’ outputs from SAS by contribution in the final model (0: non-bankrupt firms; 1: bankrupt firms)

A.1. Current liabilities ratio

A.2. Debt ratio
A.3. Return on assets ratio

A.4. Asset turnover ratio
A.5. Cash flow to short-term ratio

B. Kolmogorov-Smirnov tests’ outputs from SAS by contribution in the final model (0: non-bankrupt firms; 1: bankrupt firms)

B.1. Current liabilities ratio

<table>
<thead>
<tr>
<th>Bankruptcy</th>
<th>N</th>
<th>EDF at Maximum</th>
<th>Deviation from Mean at Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>117</td>
<td>0.504274</td>
<td>1.155625</td>
</tr>
<tr>
<td>1</td>
<td>117</td>
<td>0.290598</td>
<td>-1.155625</td>
</tr>
<tr>
<td>Total</td>
<td>234</td>
<td>0.397436</td>
<td></td>
</tr>
</tbody>
</table>

Maximum Deviation Occurred at Observation 43
Value of Current Liabilities Ratio at Maximum = 0.620983

<table>
<thead>
<tr>
<th>Kolmogorov-Smirnov Two-Sample Test (Asymptotic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KS</td>
</tr>
<tr>
<td>KSa</td>
</tr>
</tbody>
</table>
B.2. Debt ratio

Kolmogorov-Smirnov Test for Variable Debt Ratio
Classified by Variable Bankruptcy

<table>
<thead>
<tr>
<th>Bankruptcy</th>
<th>N</th>
<th>EDF at Maximum</th>
<th>Deviation from Mean at Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>117</td>
<td>0.811966</td>
<td>3.513101</td>
</tr>
<tr>
<td>1</td>
<td>117</td>
<td>0.162393</td>
<td>-3.513101</td>
</tr>
<tr>
<td>Total</td>
<td>234</td>
<td>0.487179</td>
<td></td>
</tr>
</tbody>
</table>

Maximum Deviation Occurred at Observation 85
Value of Debt Ratio at Maximum = 0.905549

Kolmogorov-Smirnov Two-Sample Test
(Asymptotic)

\[
\begin{array}{ccc}
\text{KS} & 0.324786 & D \\
\text{KSa} & 4.968275 & \text{Pr} > \text{KSa} <.0001 \\
\end{array}
\]

B.3. Return on assets ratio

Kolmogorov-Smirnov Test for Variable Return on Assets
Classified by Variable Bankruptcy

<table>
<thead>
<tr>
<th>Bankruptcy</th>
<th>N</th>
<th>EDF at Maximum</th>
<th>Deviation from Mean at Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>117</td>
<td>0.051282</td>
<td>-2.542376</td>
</tr>
<tr>
<td>1</td>
<td>117</td>
<td>0.521368</td>
<td>2.542376</td>
</tr>
<tr>
<td>Total</td>
<td>234</td>
<td>0.286325</td>
<td></td>
</tr>
</tbody>
</table>

Maximum Deviation Occurred at Observation 110
Value of Return on Assets at Maximum = -0.063415

Kolmogorov-Smirnov Two-Sample Test
(Asymptotic)

\[
\begin{array}{ccc}
\text{KS} & 0.235043 & D \\
\text{KSa} & 3.595462 & \text{Pr} > \text{KSa} <.0001 \\
\end{array}
\]
B.4. Asset turnover ratio

<table>
<thead>
<tr>
<th>Bankruptcy</th>
<th>N</th>
<th>EDF at Maximum</th>
<th>Deviation from Mean at Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>117</td>
<td>0.803419</td>
<td>2.819726</td>
</tr>
<tr>
<td>1</td>
<td>117</td>
<td>0.282051</td>
<td>-2.819726</td>
</tr>
<tr>
<td>Total</td>
<td>234</td>
<td>0.542735</td>
<td></td>
</tr>
</tbody>
</table>

Maximum Deviation Occurred at Observation 174
Value of Asset Turnover Ratio at Maximum = 1.126897

<table>
<thead>
<tr>
<th>Kolmogorov-Smirnov Two-Sample Test (Asymptotic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KS</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>KSa</td>
</tr>
<tr>
<td>Pr &gt; KSa</td>
</tr>
</tbody>
</table>

B.5. Cash flow to short-term debt ratio

<table>
<thead>
<tr>
<th>Bankruptcy</th>
<th>N</th>
<th>EDF at Maximum</th>
<th>Deviation from Mean at Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>117</td>
<td>0.401709</td>
<td>-2.357476</td>
</tr>
<tr>
<td>1</td>
<td>117</td>
<td>0.837607</td>
<td>2.357476</td>
</tr>
<tr>
<td>Total</td>
<td>234</td>
<td>0.619658</td>
<td></td>
</tr>
</tbody>
</table>

Maximum Deviation Occurred at Observation 50
Value of Cash flow to Short-term Debt at Maximum = 0.199814

<table>
<thead>
<tr>
<th>Kolmogorov-Smirnov Two-Sample Test (Asymptotic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KS</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>KSa</td>
</tr>
<tr>
<td>Pr &gt; KSa</td>
</tr>
</tbody>
</table>
C. Quantiles tables’ outputs from SAS by contribution in the final model (0: non-bankrupt firms; 1: bankrupt firms)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Current liabilities ratio</th>
<th>Estimate</th>
<th>Debt ratio</th>
<th>Estimate</th>
<th>Return on assets ratio</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>100% Max</td>
<td></td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.244755</td>
<td>7.204082</td>
<td>0.39446555</td>
</tr>
<tr>
<td>99%</td>
<td></td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.189671</td>
<td>5.319770</td>
<td>0.38878049</td>
</tr>
<tr>
<td>95%</td>
<td></td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.104991</td>
<td>4.042494</td>
<td>0.27272727</td>
</tr>
<tr>
<td>90%</td>
<td></td>
<td>1.000000</td>
<td>1.000000</td>
<td>0.964566</td>
<td>2.777778</td>
<td>0.16755171</td>
</tr>
<tr>
<td>75% Q3</td>
<td></td>
<td>0.904476</td>
<td>1.000000</td>
<td>0.838134</td>
<td>1.734607</td>
<td>0.10271903</td>
</tr>
<tr>
<td>50% Median</td>
<td></td>
<td>0.620983</td>
<td>0.809160</td>
<td>0.682734</td>
<td>1.221485</td>
<td>0.04039756</td>
</tr>
<tr>
<td>25% Q1</td>
<td></td>
<td>0.328152</td>
<td>0.561346</td>
<td>0.423581</td>
<td>0.942249</td>
<td>0.00162983</td>
</tr>
<tr>
<td>10%</td>
<td></td>
<td>0.174521</td>
<td>0.367113</td>
<td>0.261438</td>
<td>0.780370</td>
<td>-0.04257041</td>
</tr>
<tr>
<td>5%</td>
<td></td>
<td>0.103985</td>
<td>0.162171</td>
<td>0.116183</td>
<td>0.516171</td>
<td>-0.07125506</td>
</tr>
<tr>
<td>1%</td>
<td></td>
<td>0.026536</td>
<td>0.107988</td>
<td>0.035166</td>
<td>0.428571</td>
<td>-0.16836735</td>
</tr>
<tr>
<td>0% Min</td>
<td></td>
<td>0.026536</td>
<td>0.101162</td>
<td>0.019569</td>
<td>0.130735</td>
<td>-0.29428392</td>
</tr>
</tbody>
</table>
### C.4. Asset turnover ratio

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate 0</th>
<th>Estimate 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Max</td>
<td>3.62237762</td>
<td>14.708861</td>
</tr>
<tr>
<td>99%</td>
<td>3.40412620</td>
<td>10.720000</td>
</tr>
<tr>
<td>95%</td>
<td>2.02037846</td>
<td>7.000000</td>
</tr>
<tr>
<td>90%</td>
<td>1.36605657</td>
<td>5.407895</td>
</tr>
<tr>
<td>75% Q3</td>
<td>0.99057874</td>
<td>3.370000</td>
</tr>
<tr>
<td>50% Median</td>
<td>0.55170732</td>
<td>1.592870</td>
</tr>
<tr>
<td>25% Q1</td>
<td>0.27123288</td>
<td>0.959436</td>
</tr>
<tr>
<td>10%</td>
<td>0.06553723</td>
<td>0.540878</td>
</tr>
<tr>
<td>5%</td>
<td>0.02089763</td>
<td>0.410122</td>
</tr>
<tr>
<td>1%</td>
<td>0.00178891</td>
<td>0.275275</td>
</tr>
<tr>
<td>0% Min</td>
<td>0.00178891</td>
<td>0.269565</td>
</tr>
</tbody>
</table>

### C.5. Cash flow to short-term debt ratio

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate 0</th>
<th>Estimate 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Max</td>
<td>2.9556650</td>
<td>1.31034480</td>
</tr>
<tr>
<td>99%</td>
<td>2.9556650</td>
<td>1.07766990</td>
</tr>
<tr>
<td>95%</td>
<td>1.7729692</td>
<td>0.47142857</td>
</tr>
<tr>
<td>90%</td>
<td>1.1587879</td>
<td>0.31460670</td>
</tr>
<tr>
<td>75% Q3</td>
<td>0.6654344</td>
<td>0.14087760</td>
</tr>
<tr>
<td>50% Median</td>
<td>0.2861492</td>
<td>0.00504202</td>
</tr>
<tr>
<td>25% Q1</td>
<td>0.0459364</td>
<td>-0.19653179</td>
</tr>
<tr>
<td>10%</td>
<td>-0.4774194</td>
<td>-0.57575758</td>
</tr>
<tr>
<td>5%</td>
<td>-0.5507692</td>
<td>-0.68811556</td>
</tr>
<tr>
<td>1%</td>
<td>-1.5175926</td>
<td>-2.55421690</td>
</tr>
<tr>
<td>0% Min</td>
<td>-1.6397188</td>
<td>-2.55421690</td>
</tr>
</tbody>
</table>
**D. Median test’s outputs from SAS by contribution in the final model (0: non-bankrupt firms; 1: bankrupt firms)**

**D.1. Current liabilities ratio**

<table>
<thead>
<tr>
<th>Bankruptcy</th>
<th>N</th>
<th>Sum of Scores</th>
<th>Expected Under H0</th>
<th>Std Dev Under H0</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>117</td>
<td>50.0</td>
<td>58.50</td>
<td>3.832462</td>
<td>0.427350</td>
</tr>
<tr>
<td>1</td>
<td>117</td>
<td>67.0</td>
<td>58.50</td>
<td>3.832462</td>
<td>0.572650</td>
</tr>
</tbody>
</table>

Average scores were used for ties.

**Median Two-Sample Test**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistic</td>
<td>50.0000</td>
</tr>
<tr>
<td>Z</td>
<td>-2.2179</td>
</tr>
<tr>
<td>One-Sided Pr &lt; Z</td>
<td>0.0133</td>
</tr>
<tr>
<td>Two-Sided Pr &gt;</td>
<td>0.0266</td>
</tr>
</tbody>
</table>

**Median One-Way Analysis**

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.9191</td>
<td>1</td>
<td>0.0266</td>
</tr>
</tbody>
</table>

**D.2. Debt ratio**

<table>
<thead>
<tr>
<th>Bankruptcy</th>
<th>N</th>
<th>Sum of Scores</th>
<th>Expected Under H0</th>
<th>Std Dev Under H0</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>117</td>
<td>22.0</td>
<td>58.50</td>
<td>3.832462</td>
<td>0.188034</td>
</tr>
<tr>
<td>1</td>
<td>117</td>
<td>95.0</td>
<td>58.50</td>
<td>3.832462</td>
<td>0.811966</td>
</tr>
</tbody>
</table>

**Median Two-Sample Test**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistic</td>
<td>22.0000</td>
</tr>
<tr>
<td>Z</td>
<td>-9.5239</td>
</tr>
<tr>
<td>One-Sided Pr &lt; Z</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Two-Sided Pr &gt;</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

**Median One-Way Analysis**

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>90.7047</td>
<td>1</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
### D.3. Return on assets ratio

Median Scores (Number of Points Above Median) for Variable Return on Assets Classified by Variable Bankruptcy

<table>
<thead>
<tr>
<th>Bankruptcy</th>
<th>N</th>
<th>Sum of Scores</th>
<th>Expected Under H0</th>
<th>Std Dev Under H0</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>117</td>
<td>83.0</td>
<td>58.50</td>
<td>3.832462</td>
<td>0.709402</td>
</tr>
<tr>
<td>1</td>
<td>117</td>
<td>34.0</td>
<td>58.50</td>
<td>3.832462</td>
<td>0.290598</td>
</tr>
</tbody>
</table>

Average scores were used for ties.

**Median Two-Sample Test**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>83.0000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>6.3928</td>
</tr>
<tr>
<td>One-Sided Pr &gt; Z</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Two-Sided Pr &gt;</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

**Median One-Way Analysis**

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>40.8673</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF</td>
<td>1</td>
</tr>
<tr>
<td>Pr &gt; Chi-Square</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

### D.4. Asset turnover ratio

Median Scores (Number of Points Above Median) for Variable Asset Turnover Ratio Classified by Variable Bankruptcy

<table>
<thead>
<tr>
<th>Bankruptcy</th>
<th>N</th>
<th>Sum of Scores</th>
<th>Expected Under H0</th>
<th>Std Dev Under H0</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>117</td>
<td>30.0</td>
<td>58.50</td>
<td>3.832462</td>
<td>0.256410</td>
</tr>
<tr>
<td>1</td>
<td>117</td>
<td>87.0</td>
<td>58.50</td>
<td>3.832462</td>
<td>0.743590</td>
</tr>
</tbody>
</table>

Average scores were used for ties.

**Median Two-Sample Test**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>30.0000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>-7.4365</td>
</tr>
<tr>
<td>One-Sided Pr &lt; Z</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Two-Sided Pr &gt;</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

**Median One-Way Analysis**

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>55.3011</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF</td>
<td>1</td>
</tr>
<tr>
<td>Pr &gt; Chi-Square</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
### D.5. Cash flow to short-term debt

Median Scores (Number of Points Above Median) for Variable Cash flow to Short-term Debt Classified by Variable Bankruptcy

<table>
<thead>
<tr>
<th>Bankruptcy</th>
<th>N</th>
<th>Sum of Scores</th>
<th>Expected Under H0</th>
<th>Std Dev Under H0</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>117</td>
<td>80.0</td>
<td>58.50</td>
<td>3.832462</td>
<td>0.683761</td>
</tr>
<tr>
<td>1</td>
<td>117</td>
<td>37.0</td>
<td>58.50</td>
<td>3.832462</td>
<td>0.316239</td>
</tr>
</tbody>
</table>

Average scores were used for ties.

#### Median Two-Sample Test

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistic</td>
<td>80.0000</td>
</tr>
<tr>
<td>Z</td>
<td>5.6100</td>
</tr>
<tr>
<td>One-Sided Pr &gt; Z</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Two-Sided Pr &gt;</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

#### Median One-Way Analysis

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>31.4718</td>
</tr>
<tr>
<td>DF</td>
<td>1</td>
</tr>
<tr>
<td>Pr &gt; Chi-Square</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
### E. Pearson and Spearman correlation’s outputs from SAS by contribution in the final model

#### E.1. Pearson correlation

**Pearson Correlation Coefficients, N = 234**

<table>
<thead>
<tr>
<th></th>
<th>Current Liabilities Ratio</th>
<th>Debt Ratio</th>
<th>Return on Assets</th>
<th>Asset Turnover Ratio</th>
<th>Cash flow to Short-term Debt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Liabilities Ratio</td>
<td>1.00000</td>
<td>-0.01247</td>
<td>-0.05868</td>
<td>0.31784</td>
<td>-0.06487</td>
</tr>
<tr>
<td>Debt Ratio</td>
<td>0.8495</td>
<td>0.3716</td>
<td>&lt;.0001</td>
<td>0.3231</td>
<td></td>
</tr>
<tr>
<td>Return on Assets</td>
<td>-0.01247</td>
<td>1.00000</td>
<td>-0.63551</td>
<td>0.35027</td>
<td>-0.36613</td>
</tr>
<tr>
<td>Asset Turnover Ratio</td>
<td>0.3716</td>
<td>&lt;.0001</td>
<td>1.00000</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Cash flow to Short-term Debt</td>
<td>-0.06487</td>
<td>-0.36613</td>
<td>0.51114</td>
<td>-0.14009</td>
<td>1.00000</td>
</tr>
</tbody>
</table>

#### E.2. Spearman correlation

**Spearman Correlation Coefficients, N = 234**

<table>
<thead>
<tr>
<th></th>
<th>Current Liabilities Ratio</th>
<th>Debt Ratio</th>
<th>Return on Assets</th>
<th>Asset Turnover Ratio</th>
<th>Cash flow to Short-term Debt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Liabilities Ratio</td>
<td>1.00000</td>
<td>0.08756</td>
<td>0.04749</td>
<td>0.30397</td>
<td>-0.12001</td>
</tr>
<tr>
<td>Debt Ratio</td>
<td>0.1819</td>
<td>0.4697</td>
<td>&lt;.0001</td>
<td>0.0669</td>
<td></td>
</tr>
<tr>
<td>Return on Assets</td>
<td>0.08756</td>
<td>1.00000</td>
<td>-0.62799</td>
<td>0.45649</td>
<td>-0.56819</td>
</tr>
<tr>
<td>Asset Turnover Ratio</td>
<td>0.1819</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Cash flow to Short-term Debt</td>
<td>0.04749</td>
<td>0.4697</td>
<td>1.00000</td>
<td>0.0008</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

111
Appendix G – Estimates of the final model

The figures below present the rate of good classification of different estimates of the final model.

**Legend**

- **Non-bankrupt firms**
- **Bankrupt firms**
Appendix H – Model assumptions

The discriminant model has the following assumptions:\(^{34}\):

i. The predictors are not highly correlated with each other.

ii. The mean and variance of a given predictor are not correlated.

iii. The correlation between two predictors is constant across groups.

iv. The values of each predictor have a normal distribution.

In complement to these assumptions, Burns et al. (2008) listed major underlying assumptions of DA such as:

1. The observations are a random sample

2. Each predictor variable is normally distributed

3. Each of the allocations for the dependent categories in the initial classification are correctly classified

4. There must be at least two groups or categories, with each case belonging to only one group so that the groups are mutually exclusive and collectively exhaustive (all cases can be placed in a group)

5. Each group or category must be well defined, clearly differentiated from any other group(s) and natural. Putting a median split on an attitude scale is not a natural way to form groups. Partitioning quantitative variables is only justifiable if there are easily identifiable gaps at the points of division

6. For instance, three groups taking three available levels of amounts of housing loan;

7. The groups or categories should be defined before collecting the data

8. The attribute(s) used to separate the groups should discriminate quite clearly between the groups so that group or category overlap is clearly non-existent or minimal

9. Group sizes of the dependent should not be grossly different and should be at least five times the number of independent variables

\(^{34}\) IBM SPSS Statistics 19 : User guide – Discriminant Analysis Model
Appendix I – Q-Q plots

Bankruptcy status:
0: Non-bankrupt firms
1: Bankrupt firms
Appendix J – Classes of risk and scores

Different risk classes in context of the distribution scores of opposing groups

Different classes of risks as depicted in Table 3

- Non-bankrupt firms
- Bankrupt firms
- -0.8
- -0.3
- 0.4
- 2
Appendix K – Contributions

Figure A: Ratios’ contributions and scores for the specific industry studies for quantiles Q1, Q2, and Q3
Figure B: Ratios’ contributions and scores of a firm that perform good over the period of study with previous industry quantiles as benchmark.

Legend
- Q1
- Q2
- Q3
- - Non-bankrupt firm
Figure C: Ratios’ contributions and scores of a firm that perform good after a turnaround over the period of study with previous industry quantiles as benchmark.
Figure D: Ratios’ contributions and scores of a firm that went bankrupt in 2009 over the period of study with previous industry quantiles as benchmark

<table>
<thead>
<tr>
<th>Contribution 1</th>
<th>Contribution 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contribution 3</td>
<td>Contribution 4</td>
</tr>
<tr>
<td>Contribution 5</td>
<td>Scores</td>
</tr>
</tbody>
</table>

Legend: Q1, Q2, Q3, Bankrupt firm